# Developing Water and Methane Potentials for MD Simulations of Methane Clathrate Hydrate

Rory Alan John Gilmore

*Submitted in partial fulfillment*
*of the requirements of the Degree of*
*Doctor of Philosophy*

25 September 2017

School of Physics and Astronomy
Queen Mary, University of London

## Statement of originality

I, Rory Alan John Gilmore, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.



.....

Rory Alan John Gilmore
25 September 2017

## Abstract

The aim of this thesis is to develop a set of intermolecular potentials that enable the study of nucleation or decomposition of methane hydrates. The potentials are developed for water, methane, and the water-methane pair by fitting to SAPT(DFT) reference energies. The first set of potentials developed differ from recent polarisable models in that they have rank 4 ISA multipoles, rank 3 anisotropic polarisabilities, rank 3 isotropic dispersion, and anisotropic exchange-repulsion terms. These potentials are validated based on the structures and energies of small clusters and second virial coefficients.

The potentials are then significantly simplified for use in MD simulations with DL_POLY4. Simplifying the methane potential makes it difficult to fit simultaneously the global minimum dimer and a set of randomly generated dimers used as reference energies. Several methods are tested to account for polarisation in water within the limitations of DL_POLY and it is found that for MD simulations good results can be attained by increasing the charge values to match the multipole moments of a water molecule in a dielectric. Simulations are carried out for liquid water, ice Ih, and methane gas to validate the new models.

The models developed are compared in MD simulations with TIP4P/Ice and the United Atom Methane (UAM) model in simulations of sI methane clathrate; both under stable conditions and while undergoing decomposition at different temperatures. It is found that the melting behaviour differs according the methane- and water-methane interactions; the behaviour of methane under clathrate decomposition using either methane model is discussed.

## Acknowledgements

I would like to thank my supervisor and co-supervisor Dr Alston Misquitta and Prof. Martin Dove, from whom I have received an extensive level of guidance. I must also thank Chenxing Yang for advice and instruction on DL_POLY, and Min Gao and Asmi Vaghela for more of the same. I must thank Queen Mary University for funding me. I would also like to thank my family for their extraordinary patience.

# Contents

# Chapter 1

# Introduction

In the early nineteenth century it was discovered that under the right temperature and pressure conditions, water can form into “clathrate hydrates” [1, 2]; these are cages made from tens or hundreds of $H_2O$ molecules, most commonly made enclosing a “guest” molecule as illustrated in Fig 1.1. Clathrates are found in nature forming under similar temperature and pressure to common ice and are associated with geological phenomena including gas plumes on Earth and elsewhere in the solar system. The interest in clathrates containing petrochemicals, particularly methane, has increased over the last decade or so as the scale of gas clathrates became clearer and the ability to harvest methane clathrate for fuel seemed to draw closer.

At the same time, the modelling of intermolecular interactions of water is reaching a turning point. The limits of empirical potentials with simple charge and Lennard-Jones interactions seems to have been reached and there is a shift towards more complex models fitted to ab initio data. For example, the AMOEBA models for describing water [3, 4], organic molecules [5] and proteins [6] has explicit polarisability and distributed multipoles. First-principle derivations of water models have produced potentials which are able to match both the water dimer and in the bulk, which has not been possible with simpler empirical models [7]. It has also been shown that predictions of organic crystal structures are often susceptible to the accuracy of the model used and that distributed multipole models and anisotropic repulsion terms can allow for more accurate predictions. [8, 9]. Van Vleet et al. have demonstrated the functional form of potentials fitted to ab initio data may be reasonably simple, while still giving better fits to the interaction energy at a broad range of separations and giving transferable atom-atom descriptions [10].

The thesis concerns primarily the production of water, methane, and water-methane pair-potentials, fitted to a set of SAPT(DFT) reference energies. Two sets of potentials are developed: The aim of the first set is to produce "as good a model as possible". The second set is a much-simplified version of the first set and is intended to be used in large-scale MD simulations of methane clathrate. Finally, some simulations of methane clathrate are carried out allowing some the evaluation of the models and, with caution, some insights into the clathrate decomposition process.

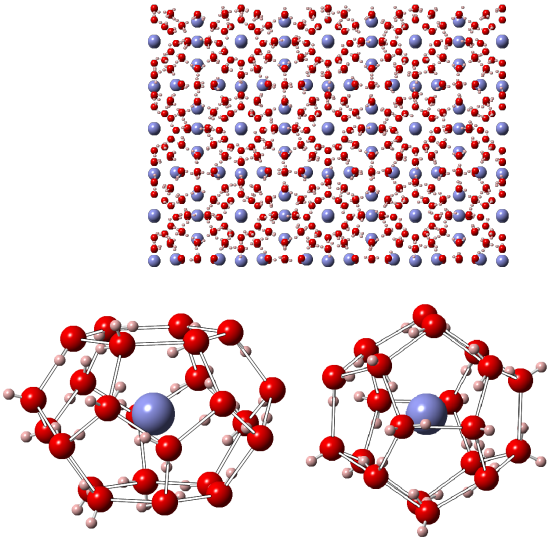


Figure 1.1: Fully-occupied sI clathrate with the guest molecule represented by the blue sphere, and the comprising $5^{12}$ and $5^{12}6^2$ cages.

## 1.1 Overview of Existing Water Potentials

Existing water potentials can be categorised in a straight-forward manner according to whether they include, for example, polarisability or anisotropy. Increasing sophistication of water models however tends to come from adding certain features to an existing model, so it can be easier to think in terms of families of water models. For the benefit of narrative this overview will take the latter approach describing some of the main water models.

### 1.1.1 TIP4P family

TIP4P models consist of a Lennard-Jones site at the position of the oxygen atom; and three point charges, one on each hydrogen atom and a third placed in the plane of the atoms "below" the oxygen site, as shown in Fig 1.2. It is sometimes noted that this four-site arrangement was present in the very earliest water model, by Bernal and Fowler [11].

There are several reparameterisations of the original TIP4P potential. The most widely applied model, often regarded as the most successful non-polarisable models, is the TIP4P/2005 model [12]. This model was fitted to match several properties with the aim to make it broadly applicable; these included the temperature of maximum density, the enthalpy of vaporisation, and the densities of liquid water and different phases of ice under appropriate pressures and temperatures.

Additionally, we make frequent comparisons to the TIP4P/Ice model [13]. This was a reparameterisation done with the intent of reproducing the phase diagram for various ice phases and for water, and is notable for its accurate melting temperature of 272.2 K.

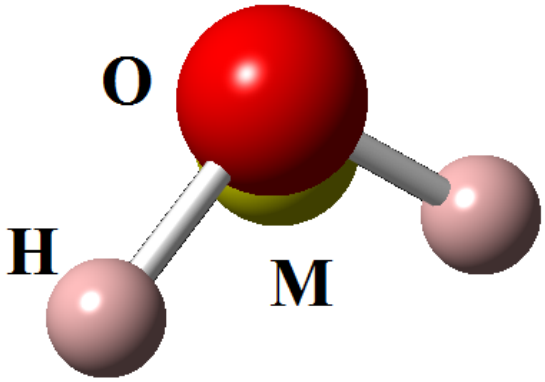



Figure 1.2: Illustration of the 4-site arrangement used for TIP4P models. M lies in the plane of the molecule, below the oxygen site. H and M are point charges, O is the site of a Lennard-Jones potential.

### 1.1.2 SAPT-5s family

The SAPT-5s potential, developed by Mas et al. (2010) [14], contains five charge sites (one on each atom and two additional sites) and three more sites which have no charge but which have repulsive walls in the form of an exponential decay prefactored by a polynomial up to $r^3$, these sites are shown in Fig 1.3. The atomic sites also have exponential walls of the same form and have dispersion coefficients up to $1/r^{10}$. Both electrostatic and dispersion terms are damped using Tang-Toennies damping.

The potential is fitted to a set of 2510 dimer energies calculated using SAPT, of these 1003 were from a fit to the earlier SAPT-pp potential [15] and were almost entirely on a grid using 8 different oxygen-oxygen separations. The remaining points were either randomly chosen between a range of separations, at random orientations with a regular set of separations, or taken from a Monte Carlo simulation of liquid water at ambient conditions.

The CC-Pol model [7] gave an increased accuracy of SAPT-5s, this used more accurate CCSD(T) reference energies for the same set of dimers and took the same form but included an extra polarisable dipole site close to the centre of mass. This led to a broad improvement but in particular made the fit quality vary less according to dimer geometry. CC-Pol has also been extended to a model CC-Pol-8s [16] with additional sets of exchange-repulsion sites and an exchange repulsion term at the location of the dipole, see Figure 1.4.

A model resembling SAPT-5s, but allowing an induced dipole at each atom, was presented in [17]. This was used as a stepping stone in producing the flexible, polarisable SAPT-5s'f potential. The introduction of flexibility to the molecule means that the number of reference energies required is greatly increased, i.e. each dimer that was used previously is replaced with a set of what is regarded as the same dimer but using different monomers. In total 239,928 grid points were used, with reference energies calculated using SAPT. From this, a flexible version of CC-Pol-8s can be produced which uses the rigid model energies with corrections taken from the SAPT-s'f potential. As the reference energies used throughout the SAPT-5s and CC-Pol models has been consistent, it is easy to compare rmse and argue that this last model

is the most accurate. It also produces the most accurate geometry for the water dimer global minimum, though the improvement for geometric parameters on the rigid CC-Pol-8s model is less that 2° for large angle $\beta$ and less than 0.01 for separation $R_{OO}$.

### 1.1.3 TTM family

The Thole-type models, so called because they relied on Thole's method to calculate induced multipoles using "smeared" charges, are among the first successful models fitted to ab initio data. The original TTM model [18] used the TIP4P charge arrangement with an induced dipole on each atom site, and an oxygen-oxygen interaction of the form $\frac{A}{R^6} + \frac{B}{R^8} - \frac{C}{R^{12}}$. Charges were fitted to reproduce molecular multipole moments up to the quadrupole and remaining parameters were fitted to match the energy along the profile of the global minimum. Improvements and expansions to the TTM model were, in chronological order; improving the size of the dataset, introducing flexibility and introducing an exponential repulsion term. These three features gave the TTM2-R (Rigid), TTM2-F (flexible) and TTM2.1-F models [19–23].

Following this, the TTM3-F [24] and TTM4-F [25] models were developed independently and concurrently. TTM3-F replaced the oxygen-oxygen interaction with a Buckingham potential and reduced the polarisability to a single site at the oxygen atom; which had zero charge. The parameters were not fitted to a set of dimers but instead to the binding energies and vibrational spectra of water clusters up to the dodecahedron. TTM4-F used an O-O interaction similar to the TTM2 model with terms up to $\frac{A}{r^{16}}$ included. Charge and dipole values were fitted to reproduce cluster energies similar to the TTM3-F model, while the O-O interaction terms were fitted to the dimer potential energy surface.

The TTM model was also used as a start-point for the development of the MB-Pol model [26]. MB-Pol uses the TTM4-F model's electrostatic and induction terms, with dispersion energy given by damped $1/r^6$ terms. The short-range energy (i.e. the repulsive wall) is given by a switching function which falls off smoothly to zero at a given cut-off; multiplied by a function dictating the monomer flexibility. Additionally, MB-Pol contains an explicit three-body short-range term fitted to around 12,000 reference energies. As a consequence, it performs

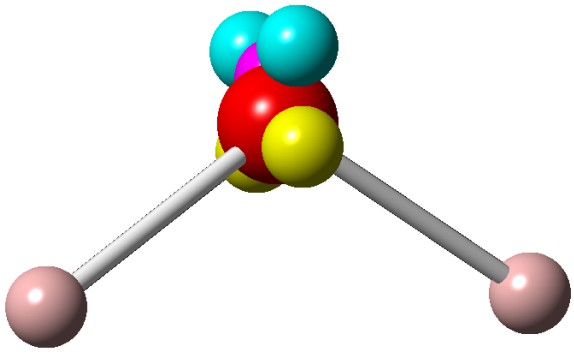


Figure 1.3: Illustration of the SAPT-5s site arrangement. Charges are located at the atomic sites and the sites marked in yellow. Repulsive walls are placed at the atomic sites and the blue and magenta sites. The displacement from the magenta site to the oxygen has been exaggerated for clarity.

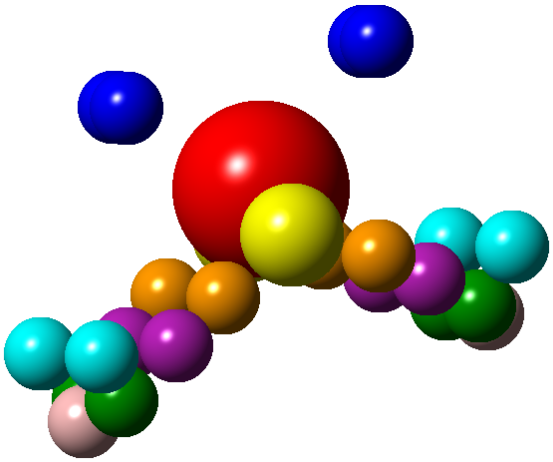


Figure 1.4: Illustration of the 25-site arrangement for CC-Pol-8s.

very well at properties such as third virial coefficient, and the energetics of water clusters, and additionally replicates liquid phase properties such as the radial distribution function well.

### 1.1.4 SCME

The SCME model [27] consists of a multipole expansion up to the hexadecapole level and polarisabilities about the centre of mass up to the quadrupole-quadrupole level about the centre of mass; and dispersion interactions between oxygen sites up to $1/R^{10}$ and an exchange-repulsion site at the oxygen atom. The electric field is preceded by a damping function at short range (this takes the form $(f_6(\beta_{ab}R_{ab}))^{1/2}$ where $f_6$ is a Tang-Toennies damping function described in Sec. 2.1.2) and a polynomial to ensure it smoothly goes to zero at long range (circa 10Å), and the exchange repulsion term increases as the density of neighbouring molecules increases (i.e as there are more molecules closer to the site). The multipole moments and polarisabilities used a combination of experiment, and calculations either at the MP2 level for multipoles moments or Hartree-Fock level for polarisabilities.

The parameters which had to be fitted were those describing the damping of the electric field, the terms of the interpolations function, and the exchange-repulsion terms including those defining the density. The electrostatic damping was made to reproduce the electric field in ice Ih and water rings up to the hexamer, the remaining terms were fitted to reproduce the MP2 energies for optimised geometries at different separations and to reproduce lattice parameters and lattice energies of ice Ih.

The SCME model was intended as an illustrative one of a functional form which avoided point charges and hence allowed for the function to be truncated after around 11Åwithout the need for Ewald summation. Although it generally performed well at reproducing the data it was designed to fit, it does not correctly predict the energetic ordering of water hexamers, and the authors suggest it may benefit from more rigorous methods of fitting parameters.

### 1.1.5 Coarse-grained Models

Whilst the general trend is for a larger number of sites, there is also a strong interested in coarse-grained models, which typically increase computational efficiency by one or two orders of magnitude. Generally it is possible to take an existing pair potential, such as TIP4P/Ew [28] or SPC/E [29], and make an isotropic pair potential which reproduces one property at the expense of another. For example, it may not be possible to reproduce in one coarse-grained model the radial density function and the tetrahedral packing arrangement; or the compressibility and the density.

The mW water model of water [30] evades these issues by including anisotropy in the form of a three-body term which penalises non-tetrahedral arrangements, based on a model for (tetrahedral) silicon. The computational efficiency associated with a coarse-grained model meant that the model could be parameterised to give an exceptionally good melting temperature at 1 atm (274.6 K), enthalpy of vapourisation, and density at ambient conditions (1.001 g $cm^{-3}$). This model contains no electrostatic interactions but instead highlights the importance of tetrahedral structure in bulk water.

Other coarse-grained models are often designed to replicate a solvent in biological applications and may by intended to represent more than one molecule. Polarisable models consisting of two or three sites have been used in place of five [31] or four [32] molecules respectively. In such cases the practical alternative would be only to include water implicitly by setting a background dielectric constant for example.

### 1.1.6 Summary of existing water models

Water models currently in use generally contain a combination of charges and dipoles, usually with four or five charge sites and sometimes using damping on the electrostatics. Some have many more sites as the source of repulsive walls, it can be taken that the practical limit of this approach has been reached. Dispersion terms of recent models typically use Tang-Toennies damping and may go up to terms of $1/r^{10}$. Some models allow for induced dipoles on (or close to) each atomic site. There is a general trend away from models designed only to match bulk properties, and into models which aim to account for both dimer energies and many-body effects more naturally through polarisation. In particular, there is increased interest in matching the structures and energy of small clusters of water. The broad range of properties and the difficulty obtaining them with empirical potentials has led to strong interest in the development of water models from first principles, which also allow polarisation to be taken into account in a natural way. Within the set of ab initio polarisable water potentials there is a tendency to rely on more and more reference points; in particular while a good rigid model has been produced with on the order of a thousand points, introducing flexibility to the model requires on the order of 100,000 points [17]. Simple rigid models have been successful at predicting bulk properties such as the phase diagram [13], while more complex polarisable rigid models such as CC-pol are able to match interaction energies for a range of dimers [7]. As discussed in Section 1.3, rigid models remain the preferred choice in studying nucleation and other phenomena through MD; and all models developed in later chapters will be rigid.

## 1.2 Overview of Existing Methane and Water-Methane Potentials

Methane has received relatively little interest in recent years, typically simulations of clathrates will use models developed in the 1990's or earlier. Methane is regarded as close to isotropic and it is most common to use single-site "united atom methane" (UAM) models interacting through a Lennard-Jones function $4\epsilon((\sigma/R)^{12} - (\sigma/R)^6)$, and the same form for a methane-oxygen interaction. These include the OPLS interaction [33] which takes parameter values from Verlet and Weis [34] fitted to reproduce the experimental pressure of liquid methane; and the model from Goodbody et al [35] which takes the same $\epsilon$ value and sets $\sigma$ to reproduce the van der Waals radius. Values for $\epsilon$ and $\sigma$ are also sometimes taken from Guillot and Guissani, although they intended a polarisation term to also be included [36]. Most common water models also have a Lennard-Jones site and the water-methane interaction is generally made using the Lorentz-Berthelot mixing rules.

Some rare exceptions include the model used by Tse et al. in a very early clathrate simulation [37], this treated methane as a single site but had separate methane-hydrogen and methane-oxygen interactions; and the model produced by Murad and Gubbins [38] which had a full set of atomic sites with partial charges.

There are a few methane-water models fitted to ab initio data: Cao et al. [39] made a single-site model with $1/R^6$ attraction and exponential repulsion fitted to MP2/cc-pVQZ reference energies. Akin-Ojo and Szalewicz [40] produced a far more complex atom-atom potential using a polynomial multiplied by an exponential decay to give the repulsive wall, coulomb interactions and a damped dispersion model with terms up to $1/R^{10}$, fitted to SAPT reference energies. Qu et al. [41] made two potentials; one using a series of polynomials of Morse exponentials and one which has an exponential wall and damped dispersion up to $1/R^{10}$.

## 1.3 Introduction to Clathrates

Although clathrates have been known about for well over a century, their interest has developed slowly. The original experiments by Davy are perhaps more notable for leading to the identification and naming of Chlorine as an element. Clathrates themselves came under focus in the early twentieth century when clathrates of methane and other small organic molecules were found to form in undersea gas pipelines leading to (sometimes fatal) plugging of pipes.

Methane clathrates can also form readily on the seabed at depths of 1-3km [42, 43], although they have been discovered at a depth of 290m [44]. It is estimated that if methane could be taken as a fuel from this source, locations of which loosely trace the coastlines of the continents, then it would provide at least twice as much energy as known oil reserves [2, 45]. Alongside the promise of energy, the large quantities of methane stored bring a threat of global warming should the methane escape into the atmosphere [2]; and previous releases are associated with the Storegga Slide event which caused the flooding of the land between Britain and mainland Europe [46].

In astrogeology, these formations are thought to be partly responsible for the atmospheres of planets and moons and may also form on comets [47]. Pluto is thought to have a decreasing quantity of noble gases in the atmosphere, possibly due to noble gases being taken in as guest

molecules [48]. Several studies have concluded that the formation of $CO_2$ clathrates on Mars is likely [49–51]; and simulation suggests that the formation of $SO_2$ and $CO_2$ clathrates can produce the relatively warm conditions of the planet in its early state [52].

Particular interest has been given to the atmosphere of Titan, which is has a level of methane far higher than that predicted according to photochemistry [53] and Choukroun et al [54] have suggested a system where plumes of methane arise from as much as sixty kilometres below the ground.

Much as there are different structures of ice dependent on pressure and temperature, there are different forms of clathrate additionally dependent on the guest molecule; and indeed in recent years two "new" forms of ice have been produced by emptying clathrates of their guest molecules [55, 56]. Generally, though, empty clathrates are thought to be unstable.

Areas of interest are the mechanism and process of nucleation, the stability under temperature/pressure and the possible clathrate types and guest molecules. The variety of clathrate types has often been treated as limited to the three most common known clathrates structures, sI, sII and sH, although recent work has digressed from this, for example Barnes and Sum's review [1] considered amorphous clathrates consisting of irregular polyhedra. Questions on stability have been the focus of much research and simulation over the past forty years, and are discussed with, and treated as, questions on bulk properties, see for example [57–59]. The process of clathrate nucleation and collapse have been a topic of molecular dynamics simulations since the first such paper appeared in 2003 [60], and decomposition in particular is increasingly the area of greatest interest.

The most commonly observed and hence most commonly simulated clathrate forms are known as sI and sII. Both structures are cubic; sI consists of two pentagonal dodecahedra $5^{12}$ cages and six larger $5^{12}6^2$ (i.e. twelve pentagonal faces and two hexagonal faces) cages in each unit cell; sII units cells consist of sixteen $5^{12}$ cages and eight $5^{12}6^4$ cages each. Common guest molecules for sI are methane and carbon dioxide, a common guest molecule for sII is hydrogen, and sII can also contain noble gases [1].

Type H clathrates were discovered in 1986 [61] and were found to have an hexagonal structure consisting of three $5^{12}$ cages, two $4^3 5^6 6^3$ cages and one $5^{12}6^8$ cage per unit cell, Xenon and hydrogen sulphide were both found to be suitable guest molecules at ambient pressure and under the correct conditions, i.e. 20 atm. and 273 K. Methane was also found it be suitable, suggesting that sH might be found in nature.

Type T clathrate [62] was discovered in 2000 and contains dimethyl ether as its guest molecule and uses four different cage types; three occupied and one unoccupied and including two new cage types; to produce a trigonal structure. The smaller, unoccupied cage appears in a lower ratio to larger cages than is found in sI, sII or sH structures, making it particularly dense. Hence it may be the case that this is the preferred structure under high pressure for a guest molecule of similar size to dimethyl ether.

There are also a number of other clathrates which, like Type T, appear only to form around certain molecules. These are typically large guest molecules and are accompanied by large cage structures, for example tert-butylamine hydrate was identified by M$^c$Mullen et al. [63] to be a cubic structure containing seventeen- and eight-sided clathrates.

Much of the current literature focuses on methane clathrate (methane prefers type sI). The leading effort to extract methane from clathrate deposits on the sea bed is believed to

come from Japan, which in March 2013 declared successful but not economically sound extraction [64], with China announcing a similar position in 2017 [65]. Studies of various gas clathrates show that the guest molecule and the clathrate type have little effect on the bulk properties, which are very similar to the bulk properties of water with the exception of thermal conductivity, which is around five times lower than ice near melting point. This discrepancy is now established as largely due to the anharmonic behaviour of the guest molecule inside its cage [66, 67].

The source of methane about which sI clathrates form naturally is still not certain, although it is often suggested that formation occurs around fluid methane, see for example Buffet et al. [43] It has been suggested that swellings on the Arctic seabed are due to the decomposition of methane clathrates which had been stored in permafrost below the seabed [68], although it is possible that the clathrate structure remains with the guest molecule dissociated, a scenario described by Jacobson et al. [69] and demonstrated in some experiments. There are also observations of large, sudden emissions of methane from permafrost [43]. Although large emissions are beyond the scope of conventional simulations, it is worth noting that Alavi et al. [70] have observed the formation of methane gas bubbles during simulations of clathrate dissociation, and the formation of methane bubbles has since been the subject simulations by Yagasaki [71] and by Bagherzadeh [72].

## 1.4 Nucleation and Decomposition of Clathrates

### 1.4.1 Overview of mechanisms

Clathrate nucleation has generally assumed to happen in one of two ways: The "labile clusters" hypothesis supposes that in supercooled water there are enough pentamer and hexamer rings that they can begin to gather into cages around a guest molecule, with the number of molecules required to form a complete cage round the molecule depending on the molecule's size and shape, and these cages then gather together, as first proposed by Christiansen and Sloan in 1994 [73]. The main alternative explanation is the "local structuring" hypothesis put forward by Radhakrishnan and Trout in 2002 [74] in response to Monte Carlo simulations. In local structuring a number of guest molecules arrange themselves in something close to the correct order through thermal fluctuations. As will be discussed, simulations generally support something closer to the local ordering hypothesis.

The preferred structure of clathrate depends on the guest molecule, in particular Sloan [45] notes that ideally the guest molecule is about nine tenths the diameter of the cage it occupies. The cages themselves obviously must bond at angles appropriate to water molecules, in particular it has been noted that the pentagonal angles found in a regular dodecahedron (as in a $5^{12}$ cage) are 108°, closer to the H-O-H angle of 104.5° than either the hexagonal or tetrahedral angles. This suggests that the most likely clathrate is one which uses the most $6^{12}$ cages, such as Jeffrey's hypothetical Type V clathrate [75]. However in forming methane clathrate, the main driving force seems to be mass-transfer: Methane has a density in clathrate two orders of magnitude higher than as a gas at ambient conditions, and has a low solubility in water. The limited availability of guest molecules may hinder formation of clathrates beyond a certain number-density of guests.

It is reasonably well-accepted that if a clathrate is melted under careful conditions, typically slightly above melting point, that it will reform much more readily than usually expected from the guest molecule in aqueous solution. This is the so-called "clathrate memory effect" [76]. Again there are two main explanations for the memory effect, with clear similarities to those explanations proposed for nucleation: Either some partial or complete cages remain in the solution, the "residual structure" hypothesis; or guest molecules remain in the solution in such a way as cages can more readily form around them, the "guest supersaturation hypothesis". A third, less common explanation is that the structure of a clathrate is imprinted onto the surface of defects or impurities, and that water molecules will consequently tend to move into place. There is no clearly correct hypothesis, and furthermore it is sometimes suggested that there is no memory effect at all [77].

### 1.4.2 MD simulations of clathrates

Early molecular dynamics (MD) simulations often used a TIP4P/2005 or other TIP4P variant and predicted correctly that methane prefers sI formation. Less common is work using polarisable models, for instance Jiang et al. [78, 79] the former of which compared thermal conductivity of the AMOEBA [3] and COS/G2 [80] polarisable models with the results of non-polarisable TIP4P/2005 and SPC/E models; and the latter of which compared temperature dependence of the COS/G2 model with that os SPC/E and TIP4P/2005. It was found that the polarisable models had thermal conductivity which better fit experimental data, but there was not a significant difference in COS/G2 and TIP4P/2005 temperature dependence. Later comparisons with DFT calculations, for instance Xue t al. (2011) [81] also found that sI clathrate was simulated reasonably well using TIP4P/2005 or SPCE/E models.

In the 2000's there was a shift in the focus of simulations towards the nucleation and decomposition processes, the first example was probably Moon et al. (2003) [60] which described formation of the small cages ($5^{12}$) found in sI and sII clathrates alongside evidence of methane structuring in the radial distribution functions; and made the frequently-repeated observation that when $5^{12}$ cages are first form they tend to be face-sharing. Later work by Hawtin et al. (2008) [82] by Walsh et al. (2009) [83] and various simulations carried out by Jacobson et al. [69, 84, 85] all provided more insight into the nucleation method. These included examples of the structure changing from sI to sII and vice versa, along with the existence of transitory phases. Hawtin et al. find that during nucleation of methane clathrate a number of $5^{12}$ cages will form and collapse in regions that coincide with a gathering of methane molecules and eventually these cages will form in a face-sharing manner, even though the sI clathrate structure methane prefers does not have any face-sharing $5^{12}$ cages. Jacobson et al.'s nucleation simulations have portrayed a chain of formations occurring, the first being described as a "blob" of guest molecules forming a cluster which is penetrated by water. Perhaps of greater interest, an amorphous crystal formed prior to the crystalline clathrate, typically consisting of irregular polyhedra of the same scale as the cages found in sI and sII (see Fig. 1.5). The time scale of the amorphous phase is on the order of microseconds and so this transitory phase should not be ignored. This paper suggests this amorphous structure was observed by Koga et al. in 2010 [86].

Work by Walsh et al. [87] using multiple MD simulations revealed that the majority of

cages present during nucleation were non-standard cages, including $4^1 5^{10} 6^2$, $4^1 5^{10} 6^3$ and $4^1 5^{10} 6^4$. Non-standard $5^{12} 6^3$ cages were also found situated between formations of sI and sII. Other work on non-standard cages was carried out by Tang et al. (2012) [88] using DFT to examine the stability of non-standard cages of between 16 and 24 $H_2O$ atoms with methane as the guest molecule. The lowest energy structure in vacuum and in water solution changed in the cases of 16 and 22 $H_2O$ molecules, but not for 18, 20 or 24 molecules. As well as the lowest energy arrangement, a number of isomers were also found. Within each of the four sets of isomers, it was found that the more adjacent pentagonal faces a cage has, the more stable it is. In aqueous solution, all of the lowest energy isomers differed from the ground state by less than 11 kJ $mol^{-1}$. This is a smaller energy difference between isomers than found in vacuum. Unlike other cages tested, it was found that in aqueous solution cages from 24 water molecules were more stable when empty than when occupied with methane. The overall results suggest that non-standard cages arise during nucleation; however the most stable cage is still found to have 20 molecules, as does the smaller cage in the sI and sII structures. Relating nucleation simulations to the earlier models of nucleation shows nothing similar to labile clusters appearing but does suggest nucleation requires a gathering of guest molecules in solution, more closely in line with local ordering. Hawtin et al. [82] ran a further simulation in which a lone methane molecule was inserted into water and after 2.5 ns found no evidence of a cage beginning to form around it.

Typically a simulation of nucleation will have on the order of 1000 molecules with simulation times on the order of hundreds of nanoseconds in the case of atomistic models or microseconds in the case of course-grained models of water (such as that developed by Jacobson et al. [89]), starting from a two-or-more-phase system in equilibrium. Some exceptions to these are Walsh et al. (2009) [83], a simulation of homogeneous nucleation which required microsecond timescales but showed nucleation from only two methane molecules in close proximity; and non-equilibrium simulation of clathrate collapse by Alavi and Ripmeester [70] which predicts significant temperature gradients upon decomposition of sI methane clathrate at a hydrate/water boundary. This paper made the additional observation that any single collapsed cage at the boundary will quickly be replaced by another and that the row bordering the water must collapse almost entirely for decomposition to occur.

There is relatively little work on clathrate formation against a backdrop of appropriate materials, or in the inclusion of impurities in a nucleation simulation. It is well-established that the inclusion of chemicals other than the intended guest molecule can affect the clathrate structure and likelihood of formation, for example mixing $SO_2$ into $CO_2$ has been shown to increase the stability of carbon dioxide clathrate, probably by occupying the smaller, $5^{12}$ cages [90]. Only very recently have these effects been tackled more seriously in MD, for example Zi et al (2016) [91] looked at the role of asphaltene in inhibiting clathrate growth. There are very clear applications in understanding what measures to take to prevent clathrate formation; or to understand the effects that impurities from the sea floor have on clathrate formation there. The need for further work was recognised by Englezos in 1993 [59] but the possible results of defects in nucleation processes are still now being speculated on, for example Pirzadeh and Kusalik (2013) [92] stated their belief that defects on the surface of ice may themselves serve as clathrate nucleation points.

One challenge associated with nucleation simulations is identifying the first presence of

clathrates. Hawtin et al. [82] searched for methane sites which were surrounded by water molecules with hydrogen bonds at the tetrahedral angle and then searched for rings to identify cage types. The past trajectories of molecules found in complete cages showed their formation. The method used by Jacobson et al., first described by them in 2009 [69], was to identify pentagonal and hexagonal rings of nearby water molecules, observing that, for instance, the standard $5^{12}6^2$ found in the sI clathrate consists of two "cups" each consisting of an hexagonal ring with a pentagonal ring connected to it at each edge. Searching for these half-cage structures also showed the clathrate formation process, in one example some half-cages were present at the end of the simulation. They also measure the largest of their "blob" formations as a measure of the gathering of guest molecules, see [93]. To investigate the variety of amorphous clathrates, Guo et al. (2011) [94] developed the "face-saturated incomplete cage analysis" or FSICA method: The water molecules and their orientations are found, then a search is made for large empty spaces which may represent the void of a clathrate cage. The nearest molecules are used to draw faces around this hollow, these molecules forming the cage. Through this method it was found that thousands of cages could exist in the amorphous structure. The same challenge does not arise in clathrate decomposition: Clathrates in MD simulations tend to melt in a row-by-row fashion, and this can generally be seen by eye as in Fig 1.6.

Decomposition of hydrates has come under stronger interest in recent years. Decomposition is regarded as a factor in geological events, but also is particularly relevant to the drive for methane extraction. Das et al (2015) [95] have studied decomposition for a variety of guest molecules and found that too large, too small, or too weakly-interacting a guest molecules will result in faster decomposition. Bagherzadeh et al. have carried out decomposition studies related to the behaviour of methane [72, 96], in which it is found that after melting guest methane tends to form into nanobubbles.

## 1.5 Structure of Thesis

From here the remaining chapters are as follows: Chapter 2 describes the process of developing pair potentials for water-water, methane-methane and water-methane. These are anisotropic potentials with rank 4 multipole models, rank 3 polarisation and dispersion models with terms up to (in the case of water) $C_{12}$. The method of producing the multipole, polarisation and dispersion models is described and details are given of fitting the short-range part of the

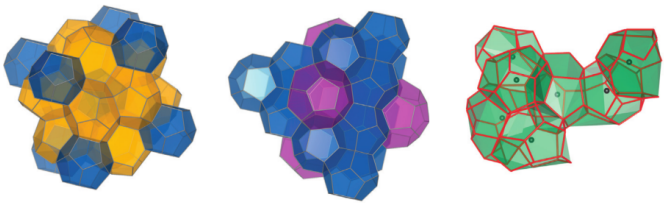


Figure 1.5: Shown left to right: the structures of sI and sII clathrates, and an example amorphous structure which appears during nucleation. $5^{12}$, $5^{12}6^2$, and $5^{12}6^4$ cages are shown in blue, yellow, and magenta respectively. The cages of the amorphous structure are irregular and non-standard. Taken from Figure 2 of Barnes and Sum 2013 [1].

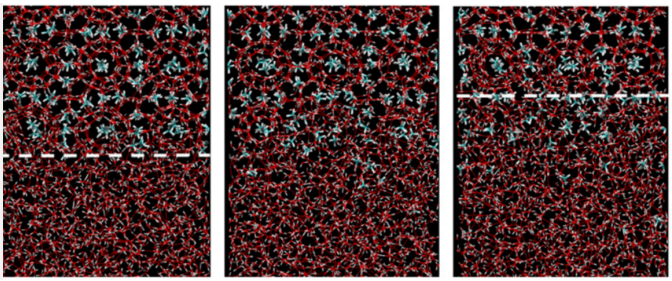


Figure 1.6: Progression of the clathrate-liquid water interface as the clathrate melts. Taken from Figure 4 of Bagherzadeh et al 2012 [97].

potentials.

In Chapter 3 these potentials are simplified so that they can be used in DL_POLY MD simulations. The dispersion model is kept untouched form the previous chapter. Several approaches are tried out to reproduce the electrostatic model and polarisability of water. The short range energy is fitted as before but with limited parameters.

Chapter 4 contains tests of the models developed in Chapters 2 and 3, including properties of small clusters and, where possible, bulk properties. This serves to validate the models but also allows final decisions to be made on the models developed in Chapter 3.

The final set of simplified potentials is tested again in Chapter 5, where the new potentials are used alongside existing models to simulate decomposition of methane clathrate. There are significant differences in the way clathrate melts using the new potentials and the commonly used set, and the reasons for this are discussed. Details of other simulations relevant to Chapter 4 are also given.

Chapter 6 contains a brief summary of results from the previous chapters, along with comments on some of the greater difficulties found along the way.

# Chapter 2

# Development of Anisotropic Pair Potentials for Water and Methane

The process of fitting the anisotropic potentials can be split broadly into two categories; molecular properties which describe the interaction at long-range, and fitting an anisotropic repulsive wall to describe the interaction at short range.

CamCASP [98, 99] is used to calculate molecular properties (i.e. distributed multipoles, polarizabilities, and dispersion coefficients), and also to produce a set of dimers and calculate SAPT(DFT) reference energies for each dimer.

The Orient [100] program is used to fit the repulsive wall; given other parts of the model and the set of reference energies. Orient is also used to calculate total energies for a given model and set of dimers. For example, when testing a water-water model energies are frequently calculated at points along the profile of the global minimum; i.e a set of dimers made by taking the global minimum dimer and altering the intermolecular separation. Subsequently, Orient is also used to perform basin-hopping searches for minima of small clusters; and to reminimise those clusters using different models.

## 2.1 Form of Reference Potential and Fitted Potential

### 2.1.1 Calculation of reference energies

The reference energies used in fitting the short-range energy, electrostatic energy and dispersion energy are SAPT(DFT) energies [101–105] with an augmented triple-zeta basis.

The interaction energy in SAPT(DFT) contains both first-order terms, i.e. those appearing at first order in perturbation theory; and higher order terms, this distinction will be convenient later. The first order terms are:

- Electrostatic energy: Energy arising classically through Coulomb's law
- Exchange-repulsion energy (or just "exchange energy"): A combination of two short-range effects; the energy required to maintain Pauli exclusion, and the smaller effect of lowering the energy by giving electrons a larger volume to move in.

and the higher-order terms:

- Induction energy: The ability of the electric field to deform a field-contributing molecule in order to lower the energy. This includes both the classical polarisation and the charge-transfer energy caused by charge-delocalisation at close range.
- Dispersion energy: A non-classical effect which can be loosely described as a correlation of electronic fluctuations between two molecules, allowing the total energy to be lowered.
- Exchange-dispersion and exchange-induction energies: These are corrections to the dispersion and induction energies when exchange effects are included.
- The Hartree-Fock estimate of the high-order induction energy.

The infinite-order interaction energy is then written as:

$$E_{\text{int}}^{(1-\infty)} = E_{\text{elst}}^{(1)} + E_{\text{exch}}^{(1)} + E_{\text{ind}}^{(2)} + E_{\text{disp}}^{(2)} + \delta_{int}^{\text{HF}} \tag{2.1}$$

Where $E_{\text{ind}}^{(2)} = E_{\text{ind,pol}}^{(2)} + E_{\text{ind,exch}}^{(2)}$ is the sum of induction and the exchange correction to induction; and $E_{\text{disp}}^{(2)} = E_{\text{disp,pol}}^{(2)} + E_{\text{disp,exch}}^{(2)}$ similarly. These five components are shown in Fig 2.1 for the global minimum profile of the water dimer.

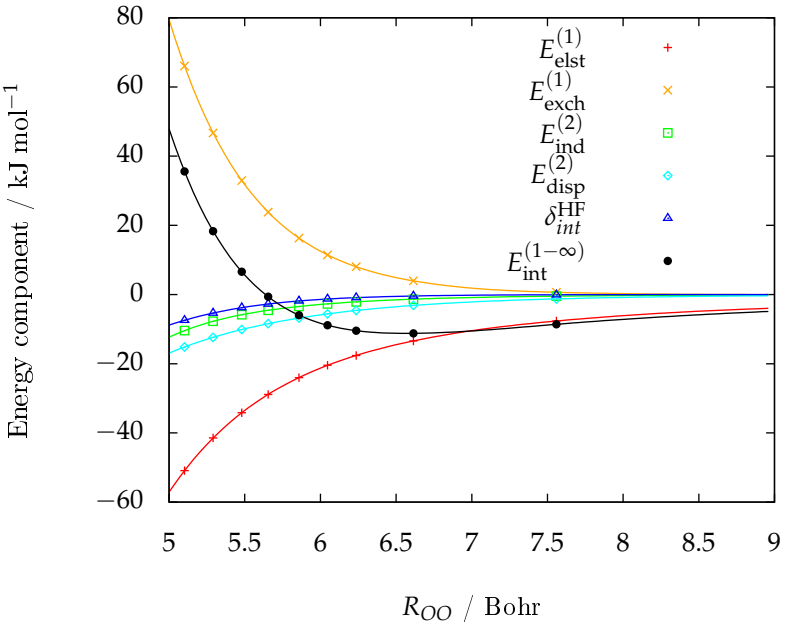



Figure 2.1: The energy contributions towards the SAPT(DFT) total interaction energy for water dimer along the global minimum profile.

The SAPT(DFT) method is based on SAPT (symmetry-adapted perturbation theory), in which the total interaction energy is split into three parts: The energies of isolated molecules in the absence of correlation effects, the contribution of electron correlation to each molecule's energy, and the intermolecular interaction. For two molecules SAPT then requires three levels of perturbation; one for the correlation energy of each molecule and one for the interaction energy. SAPT(DFT) differs from SAPT in that the energies are calculated using DFT, hence

correlation is included without the need for these first two perturbations and the calculation is simplified.

The DFT calculations use the PBE0 functional with an asymptotic correction of the form given by Tozer and Handy (1998) [106]. This correction is necessary to give accurate descriptions of exchange-correlation at long range. It requires the vertical ionisation energy, for water the value used was 0.4638 a.u. and for methane 0.4634 a.u. was used. All calculations used the ALDA+CHF kernel as described by Misquitta and Stone (2016) [98]. Energies are calculated using an augmented triple-zeta basis, including a spherical main basis set and a Cartesian auxiliary basis set. The basis set used for a molecule $A$ contains functions about $A$ and about $B$; and mid-bond functions about the point between $A$ and $B$. The mid-bond functions are included to give the dispersion energies faster convergence at long range.

The first-order terms; electrostatic energy and exchange-repulsion energy; depend on the unperturbed charge densities of the two molecules, and so these can be calculated using Kohn-Sham DFT.

The induction energy appears in perturbation theory at second-order, dependent on the overlap between the ground state of one molecule $A$ with the excited state of another molecule $B$. It is calculated from the frequency-dependent density susceptibility (FDDS) at zero frequency i.e. the change in charge density at a given point resulting from a delta-function change in charge density at another. Calculation of the FDDS uses coupled Kohn-Sham DFT [107]. The model we use also requires the polarisation energy. At second-order, the induction energy can be split into charge-transfer energy and polarisation energy using the regularised SAPT method [108]. This suppresses charge-transfer to give the second-order polarisation energy i.e. the "regularised" second-order induction energy is $E^{(2)}_{\text{ind}}(\text{Reg}) = E^{(2)}_{\text{pol}}$.

The dispersion energy is also a second-order term, it appears as the overlap between the excited states of both molecules $A$ and $B$. This is calculated from the frequency-dependent density susceptibility at imaginary frequencies, and requires time-dependent coupled Kohn-Sham DFT.

The higher-order contributions to the energy, i.e. $E^{(3-\infty)}_{\text{int}}$ are mostly induction energy and are approximated in the $\delta^{\text{HF}}_{int}$ term, i.e. induction energy in the absence of exchange-correlation effects.

### 2.1.2 Form of model

The model potential contains four terms: Electrostatic energy from distributed multipoles $V^{(1)}_{\text{elst}}$, dispersion energy $V^{(2)}_{\text{disp}}$, polarisation energy $V^{(2-\infty)}_{\text{pol}}$, and short-range energy $V^{(1)}_{\text{sr}}$. For atoms $a$ on molecule $A$ and atoms $b$ on molecule $B$ these terms can be written in that order as follows:

$$V^{(1-\infty)}_{\text{int}} = \sum_{a,b}\Big(\sum_{lk} Q^a_{lk} F_{lk}\Big) \tag{2.2}$$

$$+ \sum_{n=6,8,10,12} \Big(f_n(\beta^{\text{disp}}_{ab} R_{ab}) \frac{-C_n}{R^n_{ab}}\Big)$$

$$+ V^{(2-\infty)}_{\text{pol}}(\alpha_{lk,l'k'}, F_{lk}) \tag{2.3}$$

$$+ Ge^{-\alpha_{ab}(R_{ab}-\rho_{ab}(\Omega_{ab}))}$$

where $Q^a_{lk}$ is a multipole moment of rank $l$ given in terms of the real parts of spherical harmonics, and $F_{lk}$ is the $l$th derivative of the electric potential. In this notation index $k$ has terms $0, 1c, 1s, 2c, 2s, ..., lc, ls$, for example where $l = 1$ $k$ takes the values $0, 1c, 1s$, and for $l = 0$ $k = 0$. Here $c$ and $s$ refer to combinations of spherical harmonics into sine and cosine terms, these are formed from the regular spherical harmonics $R_{lm}$ as follows:

$$R_{l,mc} = \sqrt{\frac{1}{2}}((-1)^m R_{l,m} + R_{l,-m}) - iR_{l,ms} = \qquad \sqrt{\frac{1}{2}}((-1)^m R_{l,m} - R_{l,-m}) \tag{2.4}$$

then $Q_{lk} = \int \rho(r) R_{lk} d^3r$ for charge density $\rho(r)$. The full set of these functions is detailed in [109].

$C_n$ are coefficients of the dispersion model, fixed for each atom pair. Although in principle dispersion energy is non-additive, the non-additive contribution is expected to be relatively small for water and methane, and is neglected here.

$f_n$ is a Tang-Tonnies damping function [110] i.e.

$$f_n(\beta_{ab} R_{ab}) = 1 - \exp(-\beta_{ab} R_{ab}) \sum_{k=0}^{n} \frac{(\beta_{ab} R_{ab})^k}{k!} \tag{2.5}$$

$V^{(2-\infty)}_{\text{pol}}$ is the infinite-order polarisation energy depending on the polarisabilities and multipoles of atoms $a$ and $b$. It is possible to write polarisation energy as a pair-wise contribution preceded by a damping function; $-\frac{1}{2} f_n(\beta^{\text{Pol}}_{ab} R_{ab}) \sum_{lk,l'k'} \alpha_{lk,l'k'} F_{lk} F_{l'k'}$ where $\alpha_{lk,l'k'}$ is the rank $l$ by $l'$ polarisability and $F_{l'k'}$ is the field. However, it must be noted that the polarisation energy is not pairwise-additive and includes many-body effects which are incorporated through an iterative procedure. The polarisation energy also includes a Tang-Toennies damping function, although the damping parameters $\beta^{\text{Pol}}_{ab}$ have different values from the dispersion damping parameters.

$\rho_{ab}(\Omega_{ab})$ is the shape function for site pair $ab$ dependent on relative orientation $\Omega_{ab}$ and is given by

$$\rho_{ab}(\Omega_{ab}) = \rho^a_{ab}(\Omega_{ab}) + \rho^b_{ab}(\Omega_{ab}) \tag{2.6}$$

where $\rho^a_{ab}(\Omega_{ab}) = \sum_{l,k} \rho^a_{lk} C_{lk}(\theta_a, \phi_a)$ is the shape function for atom $a$ and $C_{lk}(\theta, \phi) = \frac{4\pi}{2l+1} Y_{l,m}(\theta, \phi)$ is a renormalised spherical harmonic term and $Y_{l,m}(\theta, \phi)$ are the usual spherical harmonics. The $C_{lk}$ that will be used are

$$C_{00} = 1 \tag{2.7}$$

$$C_{10} = \cos(\theta) \tag{2.8}$$

$$C_{11c} = \sin(\theta)\cos(\phi) \tag{2.9}$$

$$C_{11s} = \sin(\theta)\sin(\phi) \tag{2.10}$$

$$C_{20} = \frac{1}{2}(3\cos^2(\theta) - 1) \tag{2.11}$$

$$C_{22c} = \frac{\sqrt{3}}{2} \sin^2(\theta)(\cos^2(\phi) - \sin^2(\phi)) \tag{2.12}$$

In practice, Orient does not describe the isotropic part of the shape function using a separate term for each atom, but only works with a single term that is intended to represent the sum of these two terms; i.e. Orient works with a parameter

$$\rho_{00}^{ab} = \rho_{00}^{a} C_{00}(\theta_a, \phi_a) + \rho_{00}^{b} C_{00}(\theta_b, \phi_b) \tag{2.13}$$

without ever providing a value for $\rho_{00}^{a} C_{00}(\theta_a, \phi_a)$ directly.

$\alpha_{ab}$ is the hardness parameter of the short-range potential and $G$ is an energy constant chosen to be $10^{-3}$ Hartrees (2.6255 kJ mol$^{-1}$).

The multipole moments $Q_{lk}$ use the global axis system, whereas the polarisabilities $\alpha_{lk,l'k'}$ and shape functions $\rho_{ab}(\Omega_{ab})$ use local axes (i.e. defined for each atom). These are described in Appendix B.2.

## 2.2 Method of Fitting the Long-Range Terms

### 2.2.1 Fitting the Multipole Model

The multipoles used in the calculation of electrostatic energy and polarisation had been supplied from previous calculation. Two sets of multipole moments were used; the first set had been found using the DMA2 method [111] (distributed multipole analysis) with a doubly-augmented triple-zeta basis. In DMA2 a grid of points is constructed around each atom with the charge density calculated at each point. Points are then assigned a weight close to 1 at that atom's nucleus and close to zero at the boundary between it and a neighbouring atom. A multipole expansion about each site is then carried out to attempt to match this distribution. The DMA2 multipoles were found to be deficient and new multipoles found through the ISA method of Misquitta et al. (2014) [112] were used instead. This uses Lillestolen and Wheatley's [113] iterated stockholder atoms procedure, in which it is taken that the the spherically-averaged electron density of an atom $a^n$ in molecule A can be written as the total average charge density of A multiplied by the ratio of some weighting functions $w^m(r)$ which are spherically symmetric about atoms $a_m$, i.e.

$$\rho^n(r) = \rho^A(r) \frac{w^n(r)}{\sum_m w^m(r)} \tag{2.14}$$

The required weighting function $w^n$ can then be found through iterating by setting $w^n$ to be the spherical average of $\rho^n(r)$ for the next step. When the $\rho^n(r)$ converge, they are used to find the multipole moments. Misquitta et al. carried out the same procedure in basis-space, which allows for lower computational demand and gives faster convergence.

### 2.2.2 Fitting the Polarisation and Dispersion Models

The polarisation energy is calculated from local distributed polarisabilities produced using the WSM procedure [107, 114]. These local polarisabilities $\alpha_{tu}^{ab}$ are made by approximating the point-to-point polarisabilities written $\alpha_{\mathrm{PQ}}$. Here $a$ and $b$ are atomic sites, $t$ and $u$ are indices 00, 10, 10$c$ etc.; and $P$ and $Q$ refer to the positions of a point charge and a multipole respectively.

The point-to-point polarisability is calculated on a grid using the frequency-dependent density susceptibility at zero frequency and the positions $P$ and $Q$. $\alpha_{tu}^{ab}$ is adjusted; such that $\tilde{\alpha}_{\mathrm{PQ}} = \sum_{a,b} \sum_{t,u} T_{0t}^{Pa} \alpha_{tu}^{ab} T_{u0}^{bQ}$, where $T_{0t}^{Pa}$ is the interaction tensor described in [109], is approximated to match $\alpha_{\mathrm{PQ}}$. This is done first by minimising

$$S^0 = \sum_{\mathrm{P,Q}} (\tilde{\alpha}_{\mathrm{PQ}} - \alpha_{\mathrm{PQ}})^2 \tag{2.15}$$

to give local polarisability $\alpha_{tu}^{ab,0}$. This however can result in positive induction energy in cases where the grid of points for which $\alpha_{\mathrm{PQ}}$ was calculated has neglected regions between very close atoms. Subsequently, the function minimised is required to be

$$S = S^0 + \sum_{tu,t'u'} g_{tu,t'u'}(\alpha_{tu} - \alpha_{tu}^0)(\alpha_{t'u'} - \alpha_{t'u'}^0) \tag{2.16}$$

where $g_{tu,t'u'}$ are elements of some positive definite tensor; producing localised polarisabilities $\alpha_{tu}^{ab}$.

The resulting polarisabilities are used to give the classical polarisation energy. As this is a classical model a damping function as in Eq 2.5 must be included to account for penetration effects.

The dispersion energy $V_{\mathrm{disp}}^{(2)}$ is found using the local polarisabilities described above. Ignoring penetration effects it can be written as

$$V_{\mathrm{disp}}^{(2)} = -\frac{\hbar}{\pi} \sum_{a,b} T_{\alpha,\beta}^{ab} T_{\alpha',\beta'}^{ab} \int_0^\infty \alpha_{\alpha,\alpha'}^{a}(i\nu)\alpha_{\alpha,\alpha'}^{b}(i\nu)d\nu = \sum_{n=6,..,12} -\frac{C_n}{R_{ab}^n} \tag{2.17}$$

where $i\nu$ is the frequency used in calculating $\alpha_{\alpha,\alpha'}^{a}(i\nu)$ and $T_{\alpha,\beta}$ is a tensor depending on the position and orientation of molecule $A$ relative to molecule $B$, given in [109]. The coefficients $C_n$ here are anisotropic, but only isotropic models were used, made from these elements. This approach has been successful for other molecules [115]. For isotropic models, $C_n = 0$ for odd values of $n$.

With the damping function as in Eq 2.5, it is possible to damp the polarisation and dispersion models to match $E_{\mathrm{ind}}^{(2)}$ and $E_{\mathrm{disp}}^{(2)}$ using $\beta = \sqrt{2I_A} + \sqrt{2I_B}$ [114, 115], where $I_A$ and $I_B$ are the vertical ionisation energies of molecules $A$ and $B$. This is used as an estimate of the value of dispersion damping $\beta^{\mathrm{disp}}$ for each pair of molecules, which is then adjusted by hand. However, the polarisation model is not intended to match the full induction energy. Consequently, the damping parameter $\beta^{\mathrm{pol}}$ must be fitted to match $V_{\mathrm{pol}}^{(2)}$ to regularised SAPT induction energies $E_{\mathrm{pol}}^{(2)}$, as in Misquitta 2013 [108]. The full polarisation energy for the model, $V_{\mathrm{pol}}^{(2-\infty)}$, is then found by iterating until the fields and induced multipoles converge. In our case induction damping is site-dependent for water-water interactions but taken to be the same for all atom pairs in methane-methane and water-methane interactions, where the polarisation contribution to the interaction energy is smaller.

## 2.3 Method of Fitting the Short-Range Terms

The short-range terms, i.e. the repulsive walls, were fitted to a set of reference dimers using the Orient program [100]. Alongside the exchange-repulsion energy, these must try and compensate for errors in the long-range parts of the model. The fitting process takes either two or three stages and relies on the idea that each stage can give a physically realistic set of parameter values guiding the next stage of the fit. The way this is done is as follows: for a function $g(p_0, p_1, ...)$ where each free parameter $p_i$ is estimated to have a value $p_i^0$ (referred to as the "anchor value"), the function which is actually optimised is given by

$$G(p_0, p_1, ...) = g(p_0, p_1, ...) + \sum_i A_i(p_i - p_i^0)^2 \tag{2.18}$$

where $A_i$ is the "anchor strength" which must be determined by hand. In practice, only a single anchor strength $A$ was used for all parameters, except in a few cases. $g(p_0, p_1, ...)$ is the weighted RMS difference between the fitted and reference dimer energies. To carry out the optimisation, each dimer is assigned a weight $W$ according to its energy $E_{ab}$, given by

$$W(E_{ab}) = e^{-(E_{ab}-E_{\min})/E_0} \tag{2.19}$$

where $E_0$ must be chosen by the user and $E_{\min}$ is the lowest energy in the set. When fitting the anisotropic water-water potential, $E_{ab}$ was taken to be the energy being fitted at that stage; i.e. during the first order fit $E_{ab} = E^{(1)}_{\text{exch}} + E^{(1)}_{\text{pen}}$. It is not always obvious what would be a sensible value of $E_0$ using this methods, so in all later fits the total interaction energy $E^{(1-\infty)}_{\text{int}}$ was used to find $W$ instead.

### 2.3.1 Generating dimers

Short-range terms were fitted to SAPT(DFT) reference energies for sets of random but evenly distributed dimers over a chosen range of inter-molecular separations. These are produced by keeping one molecule stationary and positioning the other using a random direction vector and a random but uniformly-distributed quaternion rotation, with the separation randomly chosen over a range of $R - d_1$ to $R + d_2$ where $R$ is the van der Waals radius and $d_1$ and $d_2$ are chosen values.

### 2.3.2 Three-stage method for fitting the short-range energy

At first the short-range fit required fitting in sequence to the overlap model, the first-order short-range energy, and the infinite-order short-range energy.

#### 2.3.2.1 Fitting the overlap model

The first terms to be fitted in the short-range potential are the coefficients $K_{ab}$. This uses the charge-density overlap model [116] i.e.

$$\sum_{ab} K_{ab} \int \rho^a_{\text{el}}(r)\rho^b_{\text{el}}(r)dr = E_{\text{ovr}} \tag{2.20}$$

where $\rho^a_{\text{el}}(r)$ is the electronic density for atom $a$. The overlap model has been shown to work well for exchange energy [116], i.e. $E_{\text{ovr}} \approx E^{(1)}_{\text{exch}}$, however there is a correction to the short-range energy which is also included: the first-order penetration energy $E^{(1)}_{\text{pen}}$ results from the multipole model failing when electrons of different molecules overlap and is given by the difference between the SAPT(DFT) electrostatic energy and the electrostatic energy from the multipole model; hence the coefficients which are found are those that fit

$$\sum_{ab} K_{ab} \int \rho^a_{\text{el}}(r)\rho^b_{\text{el}}(r))dR \approx E^{(1)}_{\text{exch}} + E^{(1)}_{\text{pen}} = E^{(1)}_{\text{exch}} + E^{(1)}_{\text{elst}} - V^{(1)}_{\text{elst}} \tag{2.21}$$

where $V^{(1)}_{\text{elst}}$ is the electrostatic energy from the distributed multipole model. As the aim at this stage is only to find coefficients $K_{ab}$, the overlap model coefficients can be expected to fit well provided $E^{(1)}_{\text{exch}} + E^{(1)}_{\text{pen}}$ is proportional to $E^{(1)}_{\text{exch}}$, i.e. there is a fixed ratio of exchange-repulsion energy to penetration energy.

#### 2.3.2.2 Fitting the repulsive wall to the first-order short-range energy

The next stage is an initial fit of the decay term and shape functions to this energy; i.e. we need to find $\alpha_{ab}$ and $\rho_{ab}$ terms in

$$Ge^{-\alpha_{ab}(R_{ab}-\rho_{ab}(\Omega_{ab}))} = K_{ab}\int \rho_a^{\text{elst}}(R_{ab})\rho_b^{\text{elst}}(R_{ab})dR \tag{2.22}$$

Each parameter is now fitted with an anchor term as described above.

Note that at first-order we are considering only exchange repulsion and charge penetration, and as such the shape function for an atom depends only on that atom and not on any pair of atoms, i.e. for each atom type $m$ the shape function is given by $\rho_m = \sum_m \sum_t \rho_t^m$ where $t = 00, 10$ etc. It was found that more physically sensible results could be ensured if this was taken into account as follows:

Orient provides the parameters of the shape functions in the form $\rho_{00}^a + \rho_{00}^b$, $\rho_{10}^a$, $\rho_{20}^a$, ..., $\rho_{10}^b$, $\rho_{20}^b$, ... for every atom pair $ab$. For anisotropic terms (i.e. $\rho_{10}^a$ or higher) a single parameter was used for all atom pairs, with an initial guess taken as the value for the $aa$ atom pair. The isotropic shape function term $\rho_{00}^{ab}$ must remain as a separate parameter specific to pairs of atom-types as this is the only form which Orient accepts.

#### 2.3.2.3 Fitting the repulsive wall to the infinite-order short-range energy

The final stage absorbs remaining discrepancies between the SAPT(DFT) energies and the fitted energies into the short-range term. This step is done similarly to the previous step, with the results from the previous fit used as anchor points. Since charge-delocalisation occurs above first order; when fitting to the infinite order all terms $\rho_{lk}^a$ in pair $ab$ now depend on the atom-type of $b$ and of $a$. The shape-function terms can then only be defined as belonging to a pair of interacting atoms and not to a single site.

Due to a peculiarity with Orient the polarisation model must not be included during the fit and the polarisation energy must be subtracted from the energy to which the repulsive wall is fitted. The energy fitted to is now:

$$E_{\text{exch}}^{(1)} + E_{\text{elst}}^{(1)} + E_{\text{ind}}^{(2)} - V_{\text{pol}}^{(2-\infty)} + E_{\text{disp}}^{(2)} + \delta_{int}^{\text{HF}} \approx V_{\text{disp}}^{(2)} + V_{\text{elst}}^{(1)} + V_{\text{sr}}^{(1)} \tag{2.23}$$

### 2.3.3 Two-stage method for fitting the short-range energy

Two significant changes to the original model: The DMA multipoles were replaced with ISA multipoles; and extra parameters were included in the shape functions for atom sites in the water molecule ($\rho_{11c}$ on the hydrogens and $\rho_{22c}$ on the oxygen). In particular, the inclusion of a 22c term on the oxygen site is significant as it allows for anisotropy out of the plane of the water molecule.

With these changes made, it was found that there was no need for the initial fit to the overlap model, and instead the repulsive wall could be fitted to the first-order short-range energy with no anchor values at all.

### 2.3.4 Basin Hopping

To test the potentials, low-lying minima were found using a basin-hopping search as described by Wales and Doye (1997) [117].

In the basin-hopping method, points are selected in configuration space according to a Monte Carlo procedure. These points are minimised to their local minimum, and results from this are used to transform the potential into a series of flat "basins", this is illustrated in the one-dimensional case in Fig 2.2. When a configuration falls into a known basin then it does not need to be minimised.

The procedure used in our case is as follows: A move is made in configuration space, and if a new minimum is found, the energy of that configuration $V_{\text{new}}$ is compared to the energy of the local minimum of the previous point $V_{\text{old}}$ to decide whether to accept the move in a common Metropolis Monte Carlo format; i.e. the move is accepted for $V_{\text{new}} > V_{\text{old}}$, or for a higher energy that gives $e^{(V_{\text{old}}-V_{\text{new}})/kT}$ greater than a randomly chosen value between zero and one. If the move is rejected then a new move is made from the previous point in configuration space. In the event that a new minimum is not found, there is no test and the move is always accepted. In this way the minima are mapped out, with a preference for finding the low-lying minima.

The temperature $T$ here was set to 500 K, and the step sizes were 0.25 Bohr or 20°. 500 steps were used for each basin hopping search, this was enough to find all minima of interest and in particular the five water hexamers used to evaluate the potentials.

## 2.4 Modelling the Long Range Contributions

### 2.4.1 Modelling the electrostatic energy

Both the water model and the methane model initially used distributed multipoles made using the DMA2 method [111]. When fitting the water-methane pair potential, however, it was found that the multipoles produced this way did not give an electrostatic energy $V^{(1)}_{\text{elst}}$ matching $E^{(1)}_{\text{elst}}$ at large intermolecular separations, instead giving a ratio $V^{(1)}_{\text{elst}}/E^{(1)}_{\text{elst}} \sim 0.5$. For this reason the fits were repeated using distributed multipoles more recently developed using the DF-ISA method [112], for which this problem could be significantly reduced. Results are shown for two ISA distributed multipole models with different values of weighting parameter

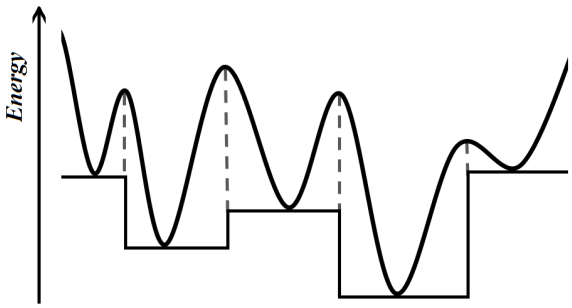


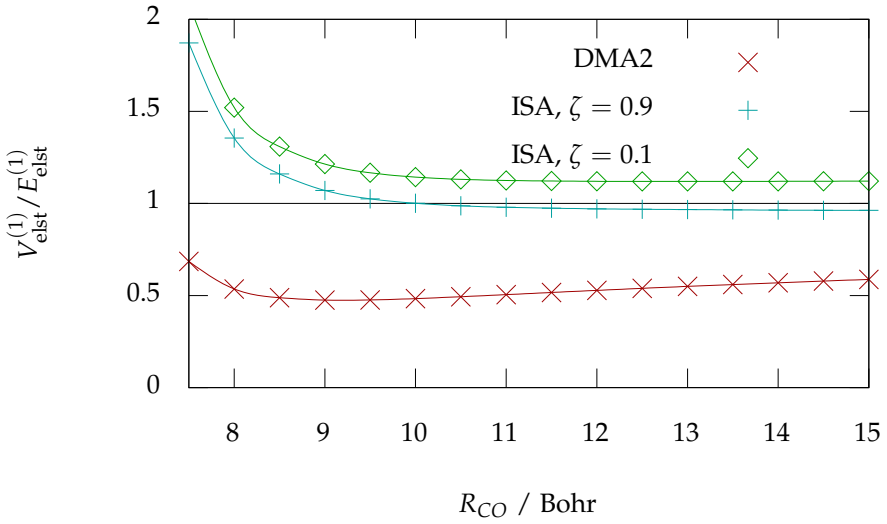

Figure 2.2: Illustration of the basin-hopping transformation of the energy surface in one dimension into a series of flat-bottomed wells. Adapted from Figure 2 in Wales and Doye 1997 [117].

$\zeta$ described in [112], which lies between 1 and 0.

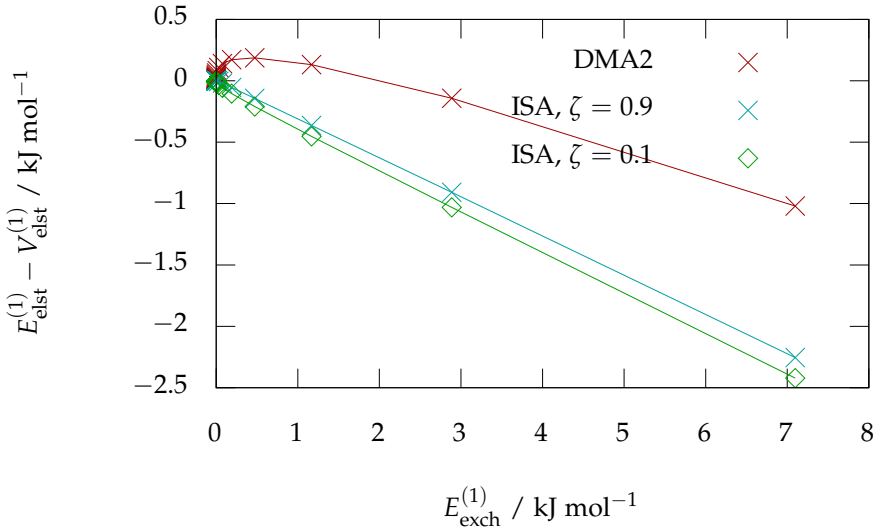
The first model tried using $\zeta = 0.9$ gave a ratio very close to 1 for the methane-water dimer at large separations $R_{CO}$, however for the water dimer it gave $V^{(1)}_{\text{elst}}/E^{(1)}_{\text{elst}} \approx 0.94$. The second set of ISA multipoles had been made with $\zeta = 0.1$ and gave $V^{(1)}_{\text{elst}}/E^{(1)}_{\text{elst}} \approx 1.1$ for the water-methane dimer and $V^{(1)}_{\text{elst}}/E^{(1)}_{\text{elst}} \approx 0.98$ for the water dimer; the water interactions are considered of greater importance and so it was this multipole model which was used. See Fig 2.3(a) for the water-methane dimer and Fig 2.4(a) for the water dimer. The same plots were made for the methane dimer however there is no clear ratio established for any multipole model in this case; this is perhaps just a consequence of the small energies involved as the electrostatic penetration energy using either ISA multipole model does fall towards zero, see Fig 2.5(a) and Fig 2.5(b).

The water and methane multipoles were also subjected to another test: That the first-order exchange energy and first-order penetration energy maintained a constant ratio i.e. $E^{(1)}_{\text{pen}} = mE^{(1)}_{\text{exch}}$ for some constant $m$. In fact, Misquitta et al. 2014 [112] have shown that this is not quite true and it is a closer approximation to say $\ln(-E^{(1)}_{\text{pen}}) = k\ln(E^{(1)}_{\text{exch}})$ for some constant $k$. However, seeking a simple proportionality aids the short-range fit in that it ensures the overlap model can be matched to $E^{(1)}_{\text{exch}} + E^{(1)}_{\text{pen}}$ and not only $E^{(1)}_{\text{exch}}$. The condition was found to hold reasonably well over most of the range of $E^{(1)}_{\text{exch}}$ for both DMA2 and ISA mutlipole models though the latter cases performed better, see Figs 2.3(b), 2.4(b) and 2.5(c) for plots of $E^{(1)}_{\text{pen}}$ against $E^{(1)}_{\text{exch}}$ for the water-methane, water, and methane dimers respectively. It is worth noting that the ratio $m$ is constant only along a given profile; i.e. constant with a change in separation only, demonstrated in the case of the water dimer (Fig 2.4(b)).

The components of the chosen ISA distributed multipole models are given in Table B.4 for water and Table B.5 for methane.

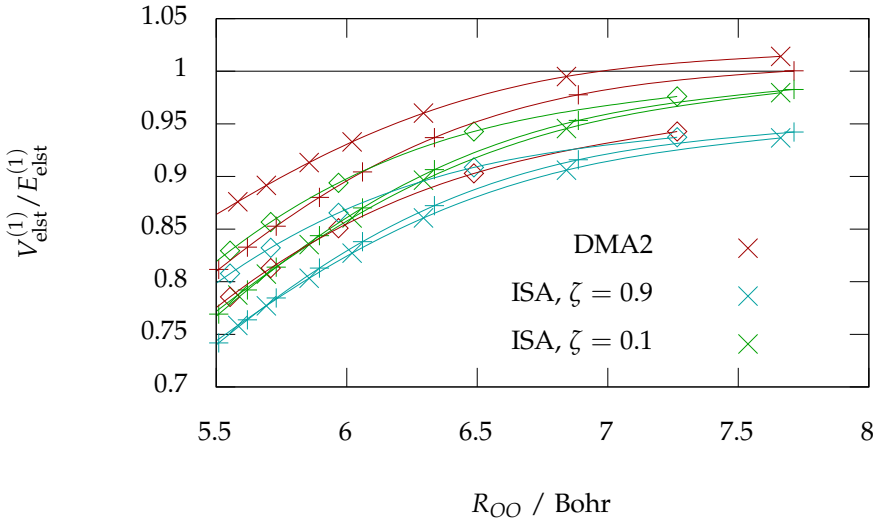



(a)

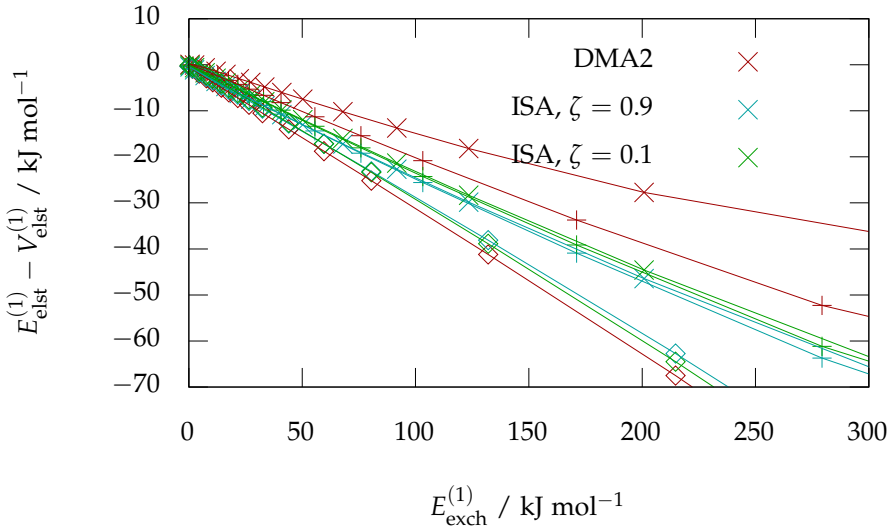



(b)

Figure 2.3: Tests of the multipole models for the water-methane dimer. Ideally the ratio $E^{(1)}_{\text{elst}}[\text{DM}]/E^{(1)}_{\text{elst}}$ should tend towards one as $R_{\text{CO}}$ increases and $E^{(1)}_{\text{exch}}/E^{(1)}_{\text{pen}}$ should be constant. The ISA model with $\zeta = 0.1$ was selected.

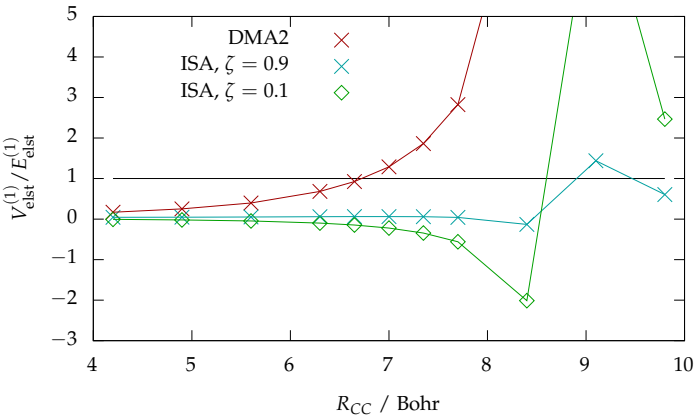



(a)

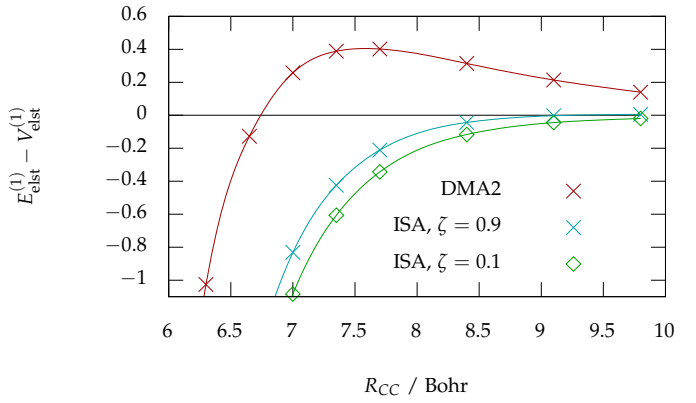



(b)

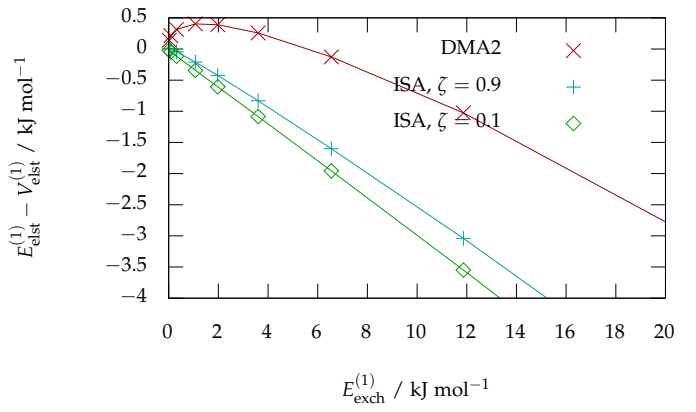
Figure 2.4: Tests of the multipole models for water. The ratio $E^{(1)}_{\text{elst}}[\text{DM}]/E^{(1)}_{\text{elst}}$ should tend towards one as $R_{\text{OO}}$ increases and $E^{(1)}_{\text{exch}}/E^{(1)}_{\text{pen}}$ should be constant for each profile; three different profiles are indicated by different marker styles. The ISA model with $\zeta = 0.1$ was selected.




(a)




(b)




(c)

Figure 2.5: Tests of the multipole models for methane. The ratio $E^{(1)}_{\mathrm{elst}}[\mathrm{DM}]/E^{(1)}_{\mathrm{elst}}$ should tend towards one as $R_{CC}$ increases. Although this does not appear to be the case, it could be that this is an artifact of the very small energies involved. $E^{(1)}_{\mathrm{exch}}/E^{(1)}_{\mathrm{pen}}$ should also be constant for a given profile. The ISA model with $\zeta = 0.1$ was selected.

### 2.4.2 Modelling the second-order energy contributions for water

The dispersion and polarisation models were produced using CamCASP as described in Sec 2.2.2. The dispersion model is isotropic and has terms up to $C_{12}$ on the O-O pair, $C_{10}$ on the O-H pair and $C_6$ only on the H-H pair. For reasons which are unclear CamCASP gives a better isotropic dispersion model when it is instead used to produce an anisotropic dispersion model and the isotropic terms are taken, see Fig 2.7. The estimate of the damping parameter from the vertical ionisation energy is 1.926 a.u., the chosen damping coefficient was slightly lower than this at $\beta^{\mathrm{pol}} = 1.89$ a.u. $C_n$ coefficients are given for this dispersion model in Table 2.1, and Fig 2.6 shows plots used to determine the damping coefficient.

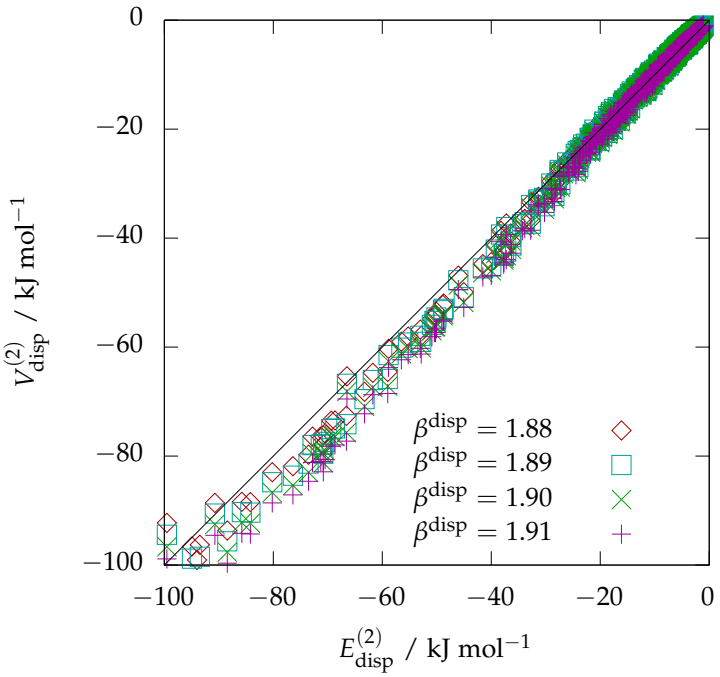
The polarisation models were anisotropic L4 models with damping dependent on the electrostatic multipole model, i.e. one set of damping parameters for the DMA model and one for the ISA model. Unlike the dispersion model, the damping parameters were chosen to be atom-pair dependent and so could not possibly be determined by eye in a single step. A scatter plot of $V^{(2)}_{\mathrm{pol}}$ against $E^{(2)}_{\mathrm{pol}}$ for the two polarisation models is given in Fig 2.8 and the two sets of damping parameters are given in Table 2.2. The full table of polarisabilities is included in Section B.5.

| Pair | $C_6$ | $C_8$ | $C_{10}$ | $C_{12}$ |
|---|---|---|---|---|
| O-O | 18.651 | 404.182 | 11886.55 | 253436.4 |
| O-H | 5.199 | 55.858 | 1081.399 | |
| H-H | 1.480 | | | |

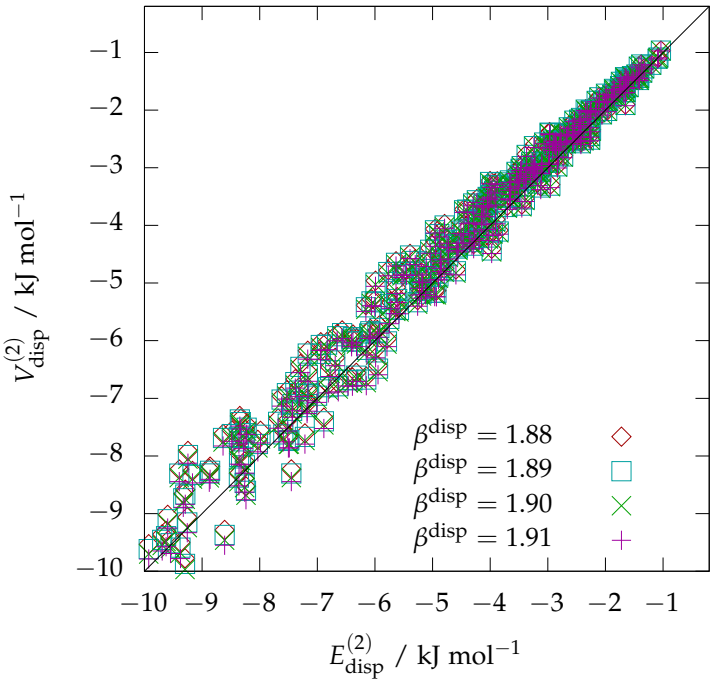
Table 2.1: Dispersion coefficients for water in a.u., taken from the anisotropic model produced by CamCASP.

| Pair | $\beta^{\mathrm{pol}}[DMA2]$ | $\beta^{\mathrm{pol}}[ISA]$ |
|---|---|---|
| O-O | 1.09 | 1.08 |
| O-H | 1.61 | 1.47 |
| H-H | 1.80 | 2.00 |

Table 2.2: Polarisation damping parameters in a.u. for use with the DMA2 and ISA multipole models.

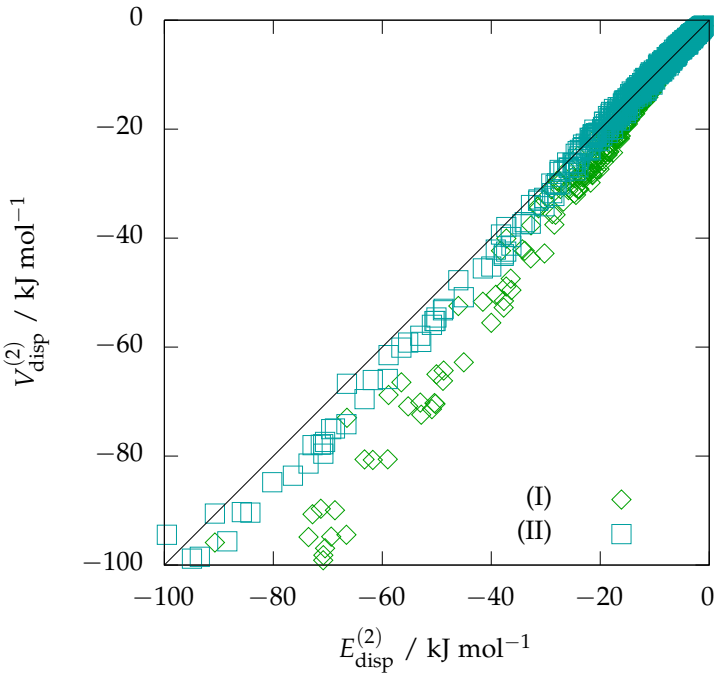



(a)

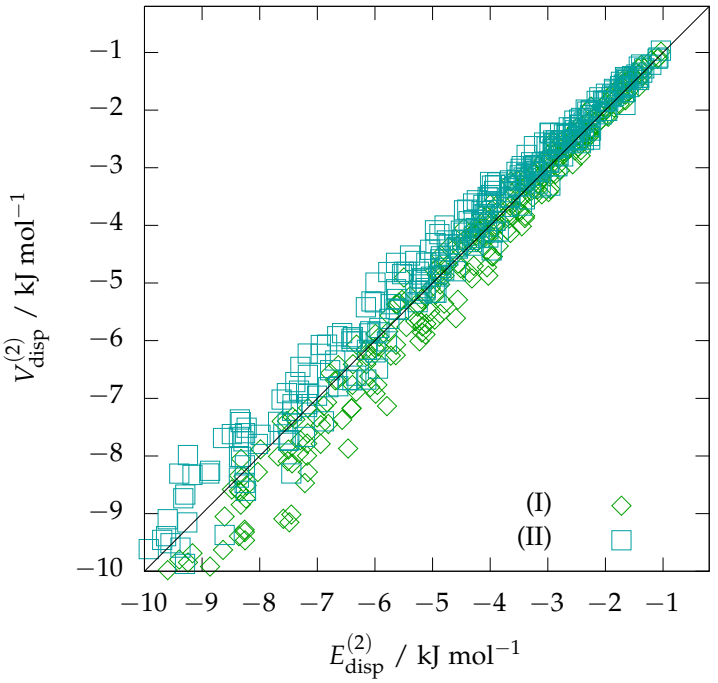



(b)

Figure 2.6: Scatter plots of $V^{(2)}_{\mathrm{disp}}$ against $E^{(2)}_{\mathrm{disp}}$ using an isotropic rank 3 dispersion model; coefficients from an anisotropic model produced by CamCASP. See also Fig 2.7.

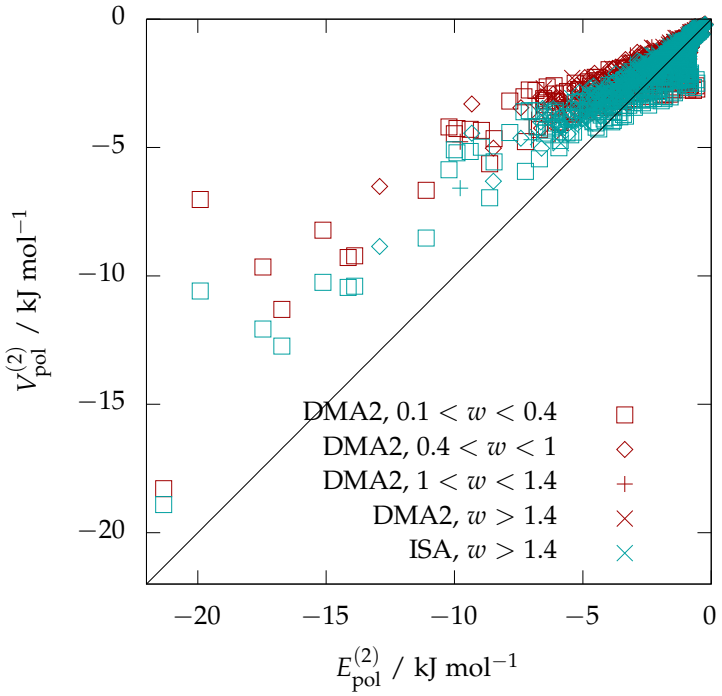



(a)

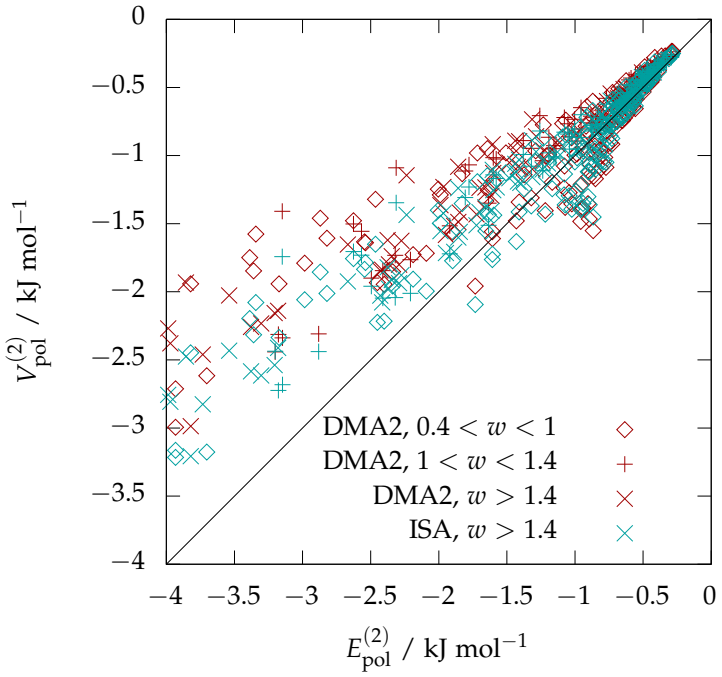



(b)

Figure 2.7: Comparison of two dispersion models, (I) an L3 isotropic model produced by CamCASP, (II) isotropic coefficients taken from an L3 anisotropic model produced by CamCASP. The majority of data points are of energy between -10 kJ mol$^{-1}$and 0 kJ mol$^{-1}$, shown in (*b*). Over either range model II gives a better fit.




(a)




(b)

Figure 2.8: Scatter plots of $V^{(2)}_{\mathrm{pol}}$ against $E^{(2)}_{\mathrm{pol}}$, i.e. the regularised second-order induction energy, using the DMA2 and ISA multipole models. Marker shapes indicate $w$, marker colours indicate model. Smaller values of $w$ indicate more positive $E^{(1-\infty)}_{\mathrm{int}}$, see Table 2.9 for details. Damping parameters are given in Table 2.2. Plots are over two ranges, most points appear in the range in the lower plot.

### 2.4.3 Modelling the second-order energy contributions for methane

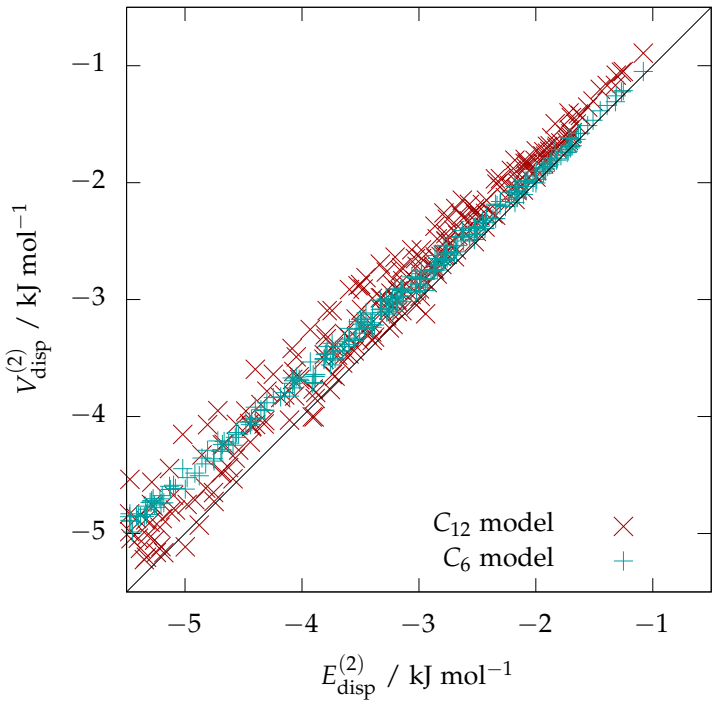
At first the dispersion model was produced using CamCASP as described in Sec 2.2.2. The dispersion model was isotropic and had terms up to $C_{12}$ on the C-C pair, $C_{10}$ on the C-H pair and $C_6$ only on the H-H pair. This model has a clear systematic error which becomes apparent at very small dispersion energies: The model appears to be offset by around $+0.16$ kJ mol$^{-1}$, see Fig 2.9. As this has greatest effect at very small dispersion energies, it cannot be corrected by adjusting the damping. For this reason the model was replaced by a $C_6$ model directly fitted to the reference $E^{(2)}_{\text{disp}}$ energies using Orient. Coefficients for both models are given in Table 2.3.

For the $C_{12}$ model setting $\beta^{\text{disp}} = 1.93$ a.u. according to methane's vertical ionisation energy seemed a reasonable choice. For the $C_6$ model it might be thought that $\beta^{\text{disp}}$ should be reduced but as points of low $E^{(1-\infty)}_{\text{int}}$ were found to typically have $E^{(2)}_{\text{disp}} > -4$ kJ mol$^{-1}$ it was decided adjustments to the damping parameter were unnecessary, see Fig 2.10.

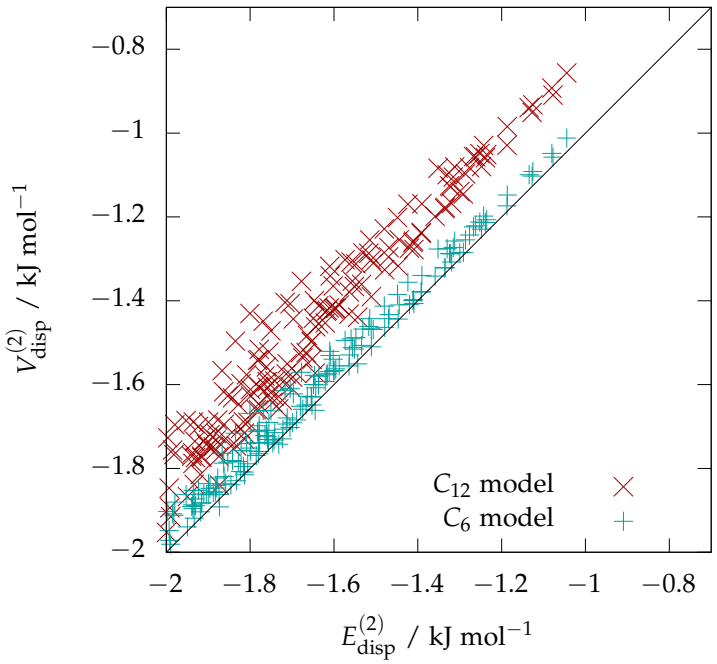
The polarisation model used was an anisotropic rank 4 model. Unlike in the water pair potential, the damping parameters were chosen to be the same for all atom pairs and so $\beta^{\text{pol}}$ was determined similarly to $\beta^{\text{disp}}$. A scatter plot of $V^{(2)}_{\text{pol}}$ against $E^{(2)}_{\text{pol}}$ for different values of $\beta^{\text{pol}}$ is given in Fig 2.11, it was decided that $\beta^{\text{pol}} = 1.93$ was a suitable choice for the methane pair potential also. The full table of polarisabilities is included in the appendix in Section B.5.

| Pair | $C_6$[Orient] | $C_6$ | $C_8$ | $C_{10}$ | $C_{12}$ |
|---|---|---|---|---|---|
| C-C | 22.079 | 23.392 | 906.143 | 51318.81 | 2023702 |
| C-H | 10.097 | 7.979 | 153.012 | 5930.860 | |
| H-H | 4.877 | 2.757 | | | |

Table 2.3: Dispersion coefficients in a.u. for two models used; the $C_6$ only model produced using Orient and the $C_{12}$ model produced by CamCASP.




(a)




(b)

Figure 2.9: Scatter plots over two energy ranges of $V^{(2)}_{\text{disp}}$ against $E^{(2)}_{\text{disp}}$ using an isotropic model with terms up to $C_{12}$, coefficients taken from an anisotropic model produced by CamCASP; and an isotropic model with $C_6$ terms only, fitted using Orient. Note the error in the $C_{12}$ model visible in 2.9(b).

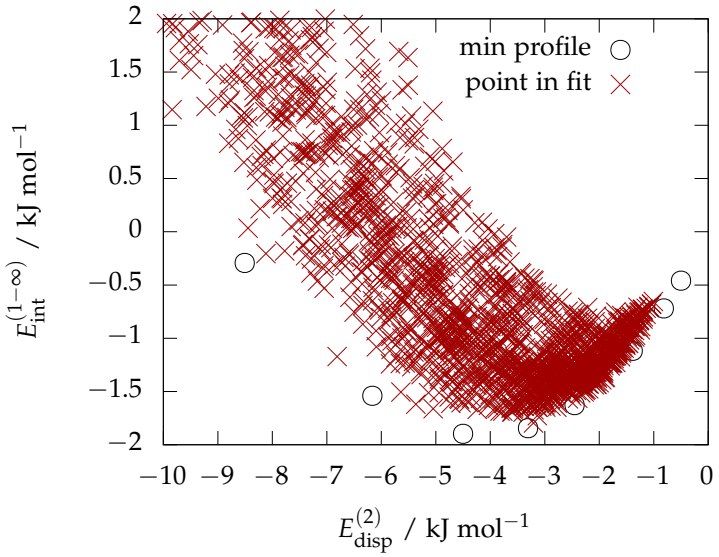



Figure 2.10: SAPT(DFT) total interaction energy against total dispersion energy for the data points used in fitting the methane pair potential. Low-energy points typically have dispersion energies between around -4 and -1 kJ mol$^{-1}$.

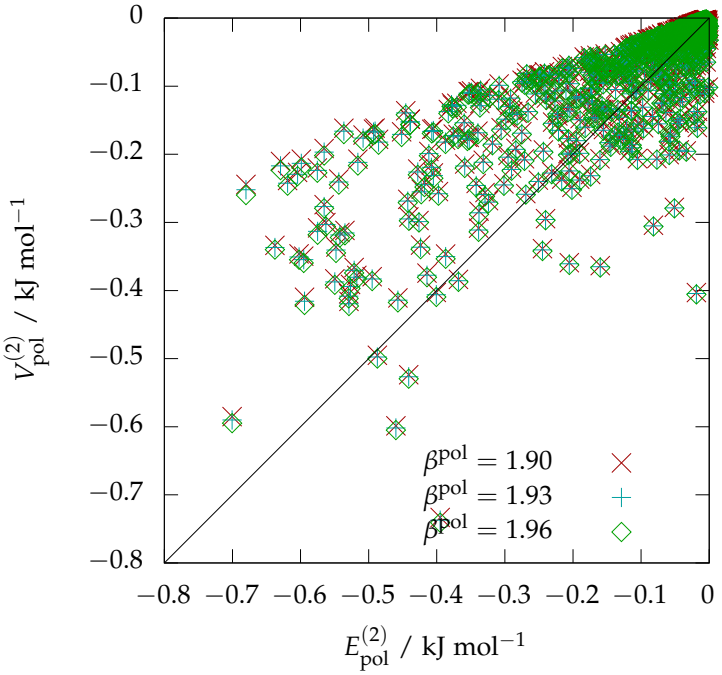



Figure 2.11: Scatter plot of $V_{\mathrm{pol}}^{(2)}$ against $E_{\mathrm{pol}}^{(2)}$ for different values of $\beta^{\mathrm{pol}}$. There is no obvious reason to deviate from $\beta^{\mathrm{pol}} = 1.93$, the value predicted by the vertical ionisation energies.

### 2.4.4 Modelling the second-order energy contributions for water-methane

The dispersion model for water-methane appears to exhibit the same error as in the methane model (Sec. 2.4.3) but with the offset reduced to around $+0.1$ kJ mol$^{-1}$, see Fig 2.12. As the total interaction energy for the water-methane interaction is larger than for the methane-methane pair interaction this systematic error was regarded as insignificant and the model was kept. The parameters for the model, which uses the isotropic terms from a rank 3 anisotropic model created with CamCASP, are given in Table 2.5. The damping parameter $\beta^{\text{disp}}$ was chosen to be 1.88 a.u., slightly lower than the 1.926 a.u. predicted using the vertical ionisation energies.

The polarisation was modelled using the polarisabilities from Sec 2.4.3 and Sec 2.4.2 with a single damping parameter $\beta^{\text{pol}}$. The choice of damping made relatively little difference to the polarisation energy although it was clear that the damping parameter had to be lower than would be predicted by the ionisation energies. $\beta^{\text{pol}} = 1.35$ a.u. was chosen, determined by plotting $V^{(2)}_{\text{pol}}$ against $E^{(2)}_{\text{pol}}$, see Fig 2.13.

| Damping | 1.80 | 1.85 | 1.88 | 1.90 | 1.93 | 1.95 |
|---|---|---|---|---|---|---|
| $w > 1$ | 0.3192 | 0.2847 | 0.2620 | 0.2657 | 0.2833 | 0.3014 |
| $0.4 < w < 1$ | 0.6696 | 0.5183 | 0.5472 | 0.6498 | 0.8481 | 0.9948 |
| $0.1 < w < 0.4$ | 0.8440 | 0.7855 | 1.6715 | 2.1168 | 2.807 | 3.2756 |
| $0.05 < w < 0.1$ | 0.8978 | 1.7449 | 3.7953 | 4.6941 | 6.0651 | 6.9900 |

Table 2.4: Table of rmse in kJ mol$^{-1}$ for different values of $\beta^{\text{disp}}$ for different regions of total interaction energy $E^{(1-\infty)}_{\text{int}}$. Larger $w$ indicates lower $E^{(1-\infty)}_{\text{int}}$, see Table 2.21 for details.

| Pair | $C_6$ | $C_8$ | $C_{10}$ | $C_{12}$ |
|---|---|---|---|---|
| C-O | 16.883 | 497.260 | 24469.86 | 668674.5 |
| C-Hw | 4.754 | 87.494 | 4100.564 | |
| H-O | 7.995 | 85.565 | 1658.857 | |
| H-Hw | 2.286 | | | |

Table 2.5: Dispersion coefficients used for water-methane interaction. “Hw” denotes hydrogen on the water molecule; “H” hydrogen on the carbon molecule.

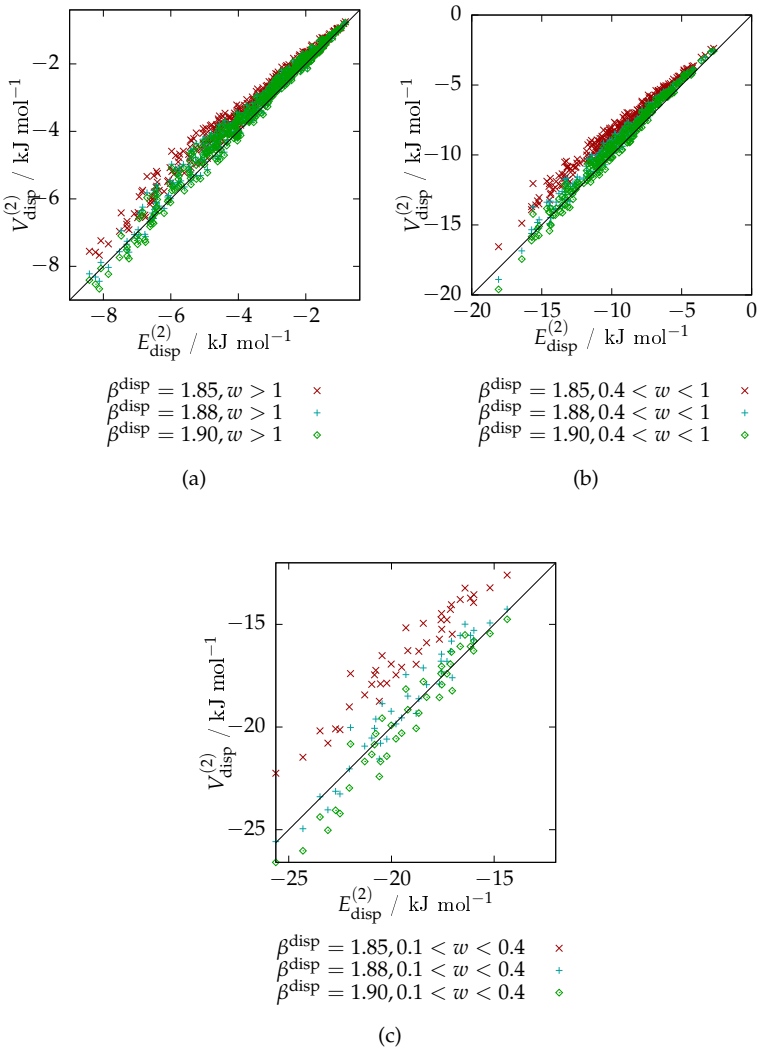



Figure 2.12: Scatter plots of $V^{(2)}_{\mathrm{disp}}$ against $E^{(2)}_{\mathrm{disp}}$ for different values of $\beta^{\mathrm{disp}}$. Plots are given according to different values of $w$, i.e. different energy ranges, see Table 2.21 for details.

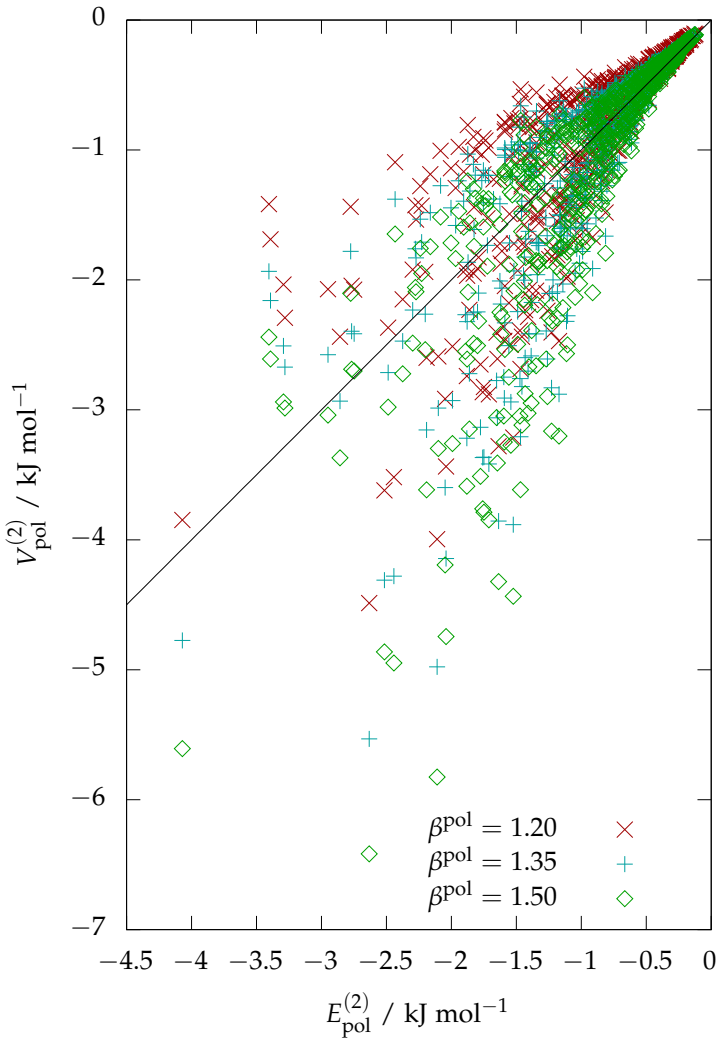



Figure 2.13: Scatter plot of $V_{\mathrm{pol}}^{(2)}$ against $E_{\mathrm{pol}}^{(2)}$ for different values of $\beta^{\mathrm{pol}}$. $\beta^{\mathrm{pol}} = 1.35$ a.u. was chosen.

## 2.5 Fitting the Short-Range Energy for Water

Fitting the short-range energy for the water potential used 2048 dimers as described in Sec 2.3.1. In earlier attempts at the fit, points along the line of the global minimum configuration with separation $R_{OO}$ changing were also used, however these were unnecessary following the introduction of the $\rho_{22c}$ term to the oxygen site and a $\rho_{11c}$ term to each hydrogen site. This allowed a more reliable description of the water molecule; in particular it is known that the oxygen is deformed out of the plane of the molecule (it's lone-pairs) and so a term allowing anisotropy out of the plane is required.

To guide parameter selection, all data points used in fitting were categorised according to the SAPT(DFT) total interaction energy. The division into different energy brackets was admittedly arbitrary and artificial, however it made it easier to focus on those dimers with the lowest energy. These categories are given in Table 2.6.

| $w = e^{-E/20\text{kJ Mol}^{-1}}$ | $E_{max}$ / kJ mol$^{-1}$ | Number of points $E_{int}$ within interval |
|---|---|---|
| $> 1.4$ | $-6.729$ | 305 |
| $> 1, < 1.4$ | 0.000 | 370 |
| $> 0.4, < 1$ | 18.326 | 662 |
| $> 0.1 < 0.4$ | 46.052 | 225 |
| $> 0.001 < 0.1$ | 138.155 | 139 |

Table 2.6: Method of splitting the reference points according to SAPT(DFT) total interaction energy, for the sake of fitting the short-range energy at first- and infinite-order.

#### 2.5.0.1 Fitting the repulsive wall to first-order

The first-order energy does not require a preliminary fit to the overlap model or to be anchored to any value, hence all that is required is a good choice of $E_0$. A value of 15 kJ mol$^{-1}$ was picked for $E_0$ with the intent of finding some compromise between low energies and the repulsive wall, see Table 2.7 and Fig 2.14. The resulting potential is given in Table 2.8.

| $E_0$ /kJ mol$^{-1}$ | 5 | 10 | 15 | 20 | 25 |
|---|---|---|---|---|---|
| $w > 1.4$ | 0.2703 | 0.2741 | 0.3461 | 0.4031 | 0.4477 |
| $> 1, < 1.4$ | 0.4541 | 0.3164 | 0.3175 | 0.3435 | 0.3708 |
| $> 0.4, < 1$ | 0.8445 | 0.4873 | 0.3665 | 0.3397 | 0.3484 |
| $> 0.1 < 0.4$ | 6.2491 | 4.1042 | 3.2823 | 2.6603 | 2.0984 |
| $> 0.001 < 0.1$ | 15.7085 | 11.1326 | 9.3250 | 7.9654 | 6.7390 |

Table 2.7: rmse in kJ mol$^{-1}$ from varying the weighting parameter $E_0$ when fitting to the first-order energy. $w$ as in Table 2.6.

#### 2.5.0.2 Fitting the repulsive wall to infinite-order

The potential from Table 2.8 was then used to give a set of anchor points when fitting to the infinite-order short-range energy. It was found weaker anchors make little difference to rmse or to the fit along the global minimum past anchors strengths of $A = 0.01$ so this value was chosen, see Table 2.9 and Figures 2.17 and 2.18. The Boltzmann weighting parameter had little effect on the scatter plot for energies below around 50 kJ mol$^{-1}$ when $E_0 > 10$ kJ mol$^{-1}$,

| Site (pair) | $l$ ($l_a$,$l_b$) | $\rho$ | $\alpha$ |
|---|---|---|---|
| O,O | 00,00 | 5.903821 | 1.780629 |
| O,H | 00,00 | 4.704752 | 1.987617 |
| H,H | 00,00 | 3.688549 | 1.899784 |
| O | 10 | 0.015407 | |
| | 20 | 0.017781 | |
| | 22c | $-0.136232$ | |
| H | 10 | $-0.119080$ | |
| | 11c | 0.031511 | |

Table 2.8: Parameters (in a.u.) in the repulsive wall when fitted to first-order, i.e. when fitted to match the sum of the exchange energy and the charge-penetration energy.

see Table 2.10 and Figure 2.19. Setting $E_0 = 10$ kJ mol$^{-1}$ gave the best fit along the global minimum profile (Fig 2.20), however setting $E_0 = 15$ kJ mol$^{-1}$ also gave a good fit and gave a lower rmse in most of the energy range. Ultimately, $E_0 = 10$ kJ mol$^{-1}$ was chosen because it came closest to getting the correct ordering of the two lowest-lying hexamers, the Cage and the Prism, see Table 2.11. The final set of parameters for the short-range potential are given in Table 2.12.

The final shape functions are shown in Figures 2.15 and 2.16. These imply the benefits of including the 22c terms on the oxygen atom, which tends to stretch in the "y" direction, perpendicular to the molecule. Note also that the shape bulges out slightly in the negative z-direction, towards the bisector of the hydrogen sites. Each hydrogen atom is drawn slightly towards the other but largely shifted towards the oxygen atom, there is very little anisotropy in the XY-plane. The 11c terms on the hydrogen atom are very small perhaps not required.

| Energy Range/kJ mol$^{-1}$ | $A = 0.1$ | $A = 0.01$ | $A = 0.001$ | no anchors |
|---|---|---|---|---|
| $w > 1.4$ | 0.2430 | 0.2400 | 0.2399 | 0.2399 |
| $> 1, < 1.4$ | 0.2329 | 0.2194 | 0.2173 | 0.2170 |
| $> 0.4, < 1$ | 0.2772 | 0.2779 | 0.2814 | 0.2816 |
| $> 0.1, < 0.4$ | 1.2648 | 1.2136 | 1.2087 | 1.2082 |
| $> 0.001 < 0.1$ | 4.3959 | 4.5284 | 4.5670 | 4.5724 |

Table 2.9: rmse in kJ mol$^{-1}$ from varying the anchors strength of $\alpha$'s and $\rho$'s when fitting the water potential to the infinite-order energy, using weighting parameter $E_0 = 15$kJ mol$^{-1}$.

| $E_0$ /kJ mol$^{-1}$ | 5 | 10 | 15 | 20 |
|---|---|---|---|---|
| $w > 1.4$ | 0.1626 | 0.1986 | 0.2399 | 0.2743 |
| $> 1, < 1.4$ | 0.2006 | 0.2019 | 0.2194 | 0.2389 |
| $> 0.4, < 1$ | 0.4401 | 0.3028 | 0.2799 | 0.2815 |
| $> 0.1, < 0.4$ | 2.8953 | 1.7917 | 1.2136 | 0.9241 |
| $> 0.001 < 0.1$ | 9.7829 | 6.2581 | 4.5284 | 3.6441 |

Table 2.10: rmse in kJ mol$^{-1}$ from varying the weighting parameter $E_0$ when fitting to the infinite-order energy, for anchor strengths set to 0.01.

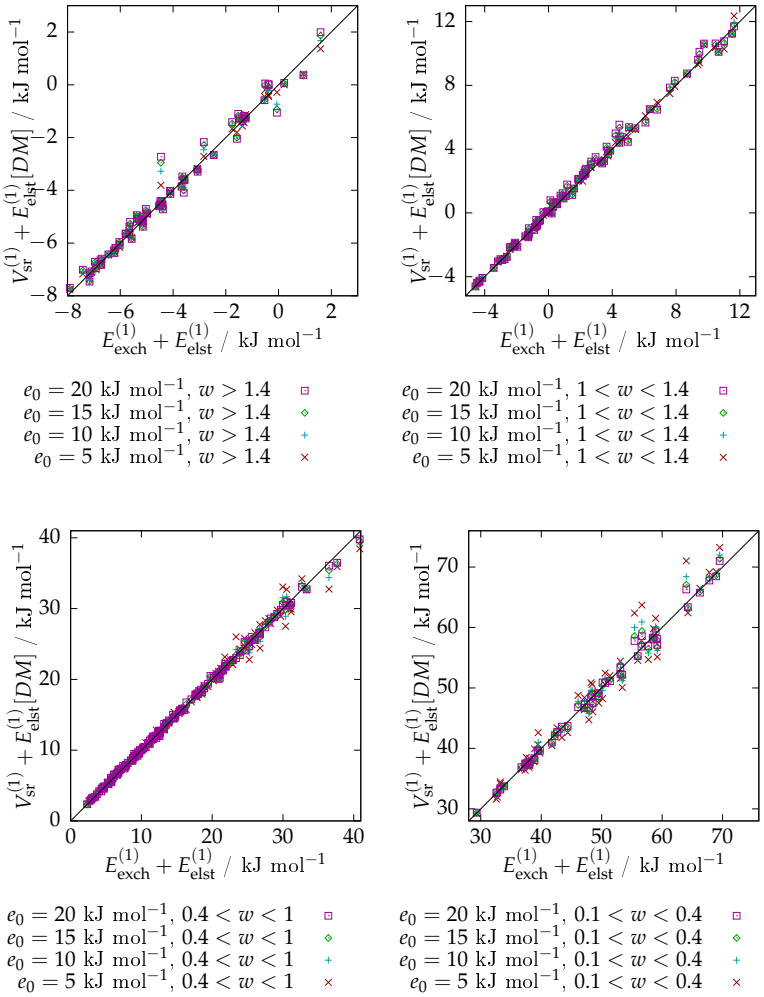



Figure 2.14: First-order energy $V_{sr}^{(1)} + V_{elst}^{(1)}$ when varying the weighting parameter $E_0$ when fitting the repulsive wall $V_{sr}^{(1)}$ to the sum of first-order exchange and charge-penetration energies $E_{exch}^{(1)} + E_{pen}^{(1)}$. Plots are for different values of $w$, i.e. different energy ranges, see Table 2.21.

| | Prism | Cage | Book(1) | Bag | Ring |
|---|---|---|---|---|---|
| $E_0 = 10$ kJ mol$^{-1}$ | 0.00 | −0.157 | 1.613 | 3.965 | 5.630 |
| $E_0 = 15$ kJ mol$^{-1}$ | 0.00 | −0.243 | 0.933 | 3.250 | 4.76 |
| CCSD(T)$^a$ | 0.00 | 0.586 | 3.222 | - | 6.778 |
| CCSD(T)$^b$ | 0.00 | 0.88 | 2.97 | 6.57 | 7.66 |

Table 2.11: Relative Binding Energies of Water Hexamers in kJ mol$^{-1}$. *a* from [118], CCSD(T) using aug-cc-pVTZ dataset optimised at MP2/aug-cc-pVTZ. *b* from [119]

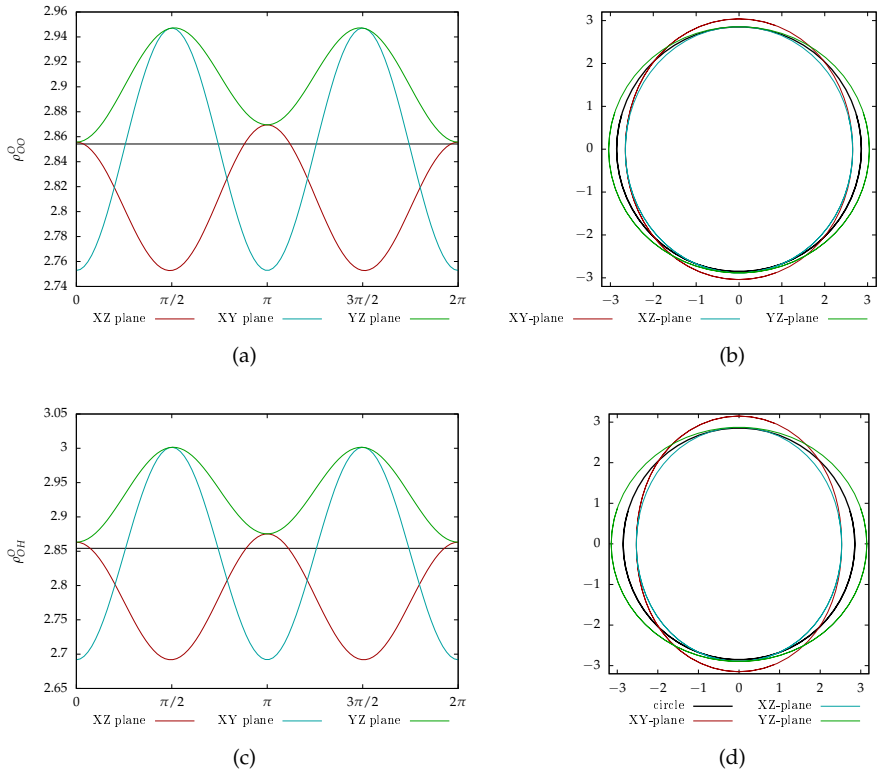



Figure 2.15: Plots of the final shape functions for the oxygen atom in the *OO* ((a) and (b)) and *OH* ((c) and (d)) pairs, the latter of which shows slightly more anisotropy. The black line indicates the spherical average. The XZ and YZ planes are measured from the Z axis, the XY plane is measured from the X axis. For parametric plots, the anisotropic terms have been increased by a factor of 2 to be seen more clearly. The Z axis points toward the hydrogen bisector, the X-axis points in the direction between hydrogens, axis details are given in Appendix B.2.

| Site (pair) | $l_a$,$l_b$ | $\rho$ | $\alpha$ |
|---|---|---|---|
| O,O | 00,00 | 5.725084 | 1.92296 |
| | 10,- | 0.000441 | |
| | 20,- | 0.009604 | |
| | 22c,- | $-0.100775$ | |
| O,H | 00,00 | 4.775246 | 1.878702 |
| | 10,- | 0.004987 | |
| | -, 10 | $-0.277651$ | |
| | -, 11c | 0.033224 | |
| | 20,- | 0.015348 | |
| | 22c,- | $-0.178707$ | |
| H,H | 00,00 | 3.677247 | 1.932032 |
| | 10, - | $-0.123442$ | |
| | 11c,- | $-0.002921$ | |

Table 2.12: Parameters in the repulsive wall when fitted to the infinite-order short-range energy.

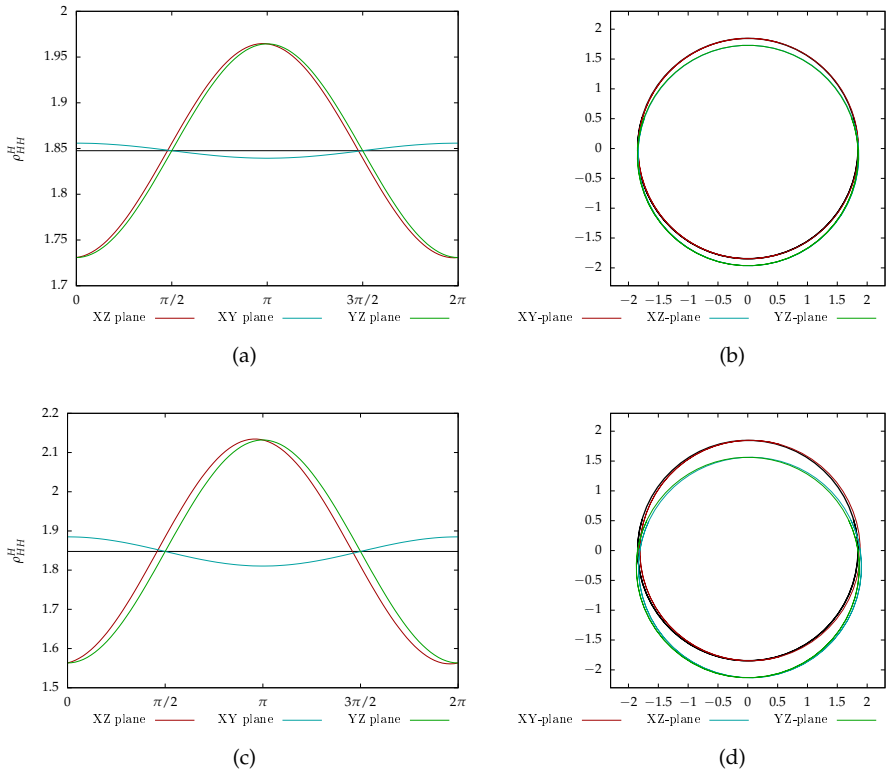



Figure 2.16: Plots of the final shape functions for the hydrogen atom in the *OO* and *OH* pairs, the latter of which shows much more anisotropy. The black line indicates the spherical average. The XZ and YZ planes are measured from the Z axis, the XY plane is measured from the X axis. The Z-axis points along the OH line, the X-axis is perpendicular to this in the plane of the molecule. Axis details are given in Appendix B.2.

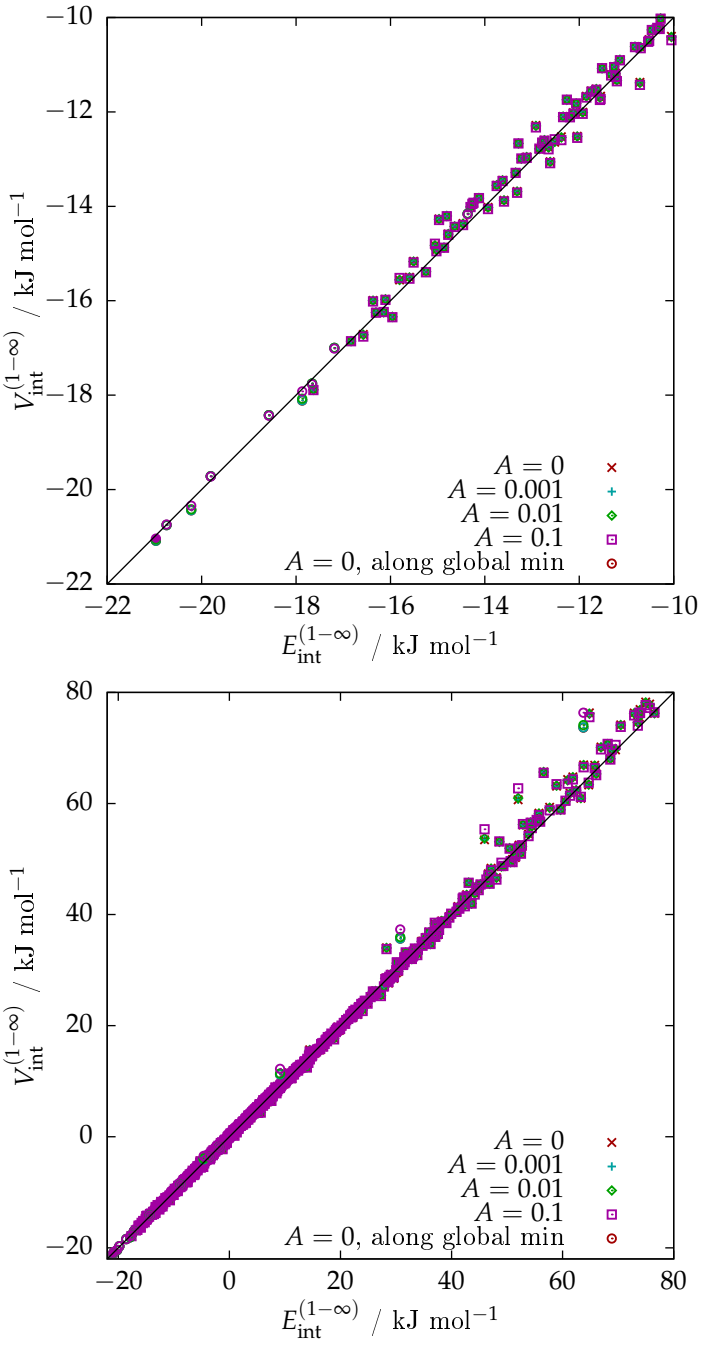



Figure 2.17: Total interaction energies varying the anchor strength $A$ of $\alpha$'s and $\rho$'s when fitting the water potential to the infinite-order energy, using weighting parameter $E_0 = 15$kJ mol$^{-1}$. Plots are over different energy ranges. Circular markers indicate points along the profile of the global minimum.

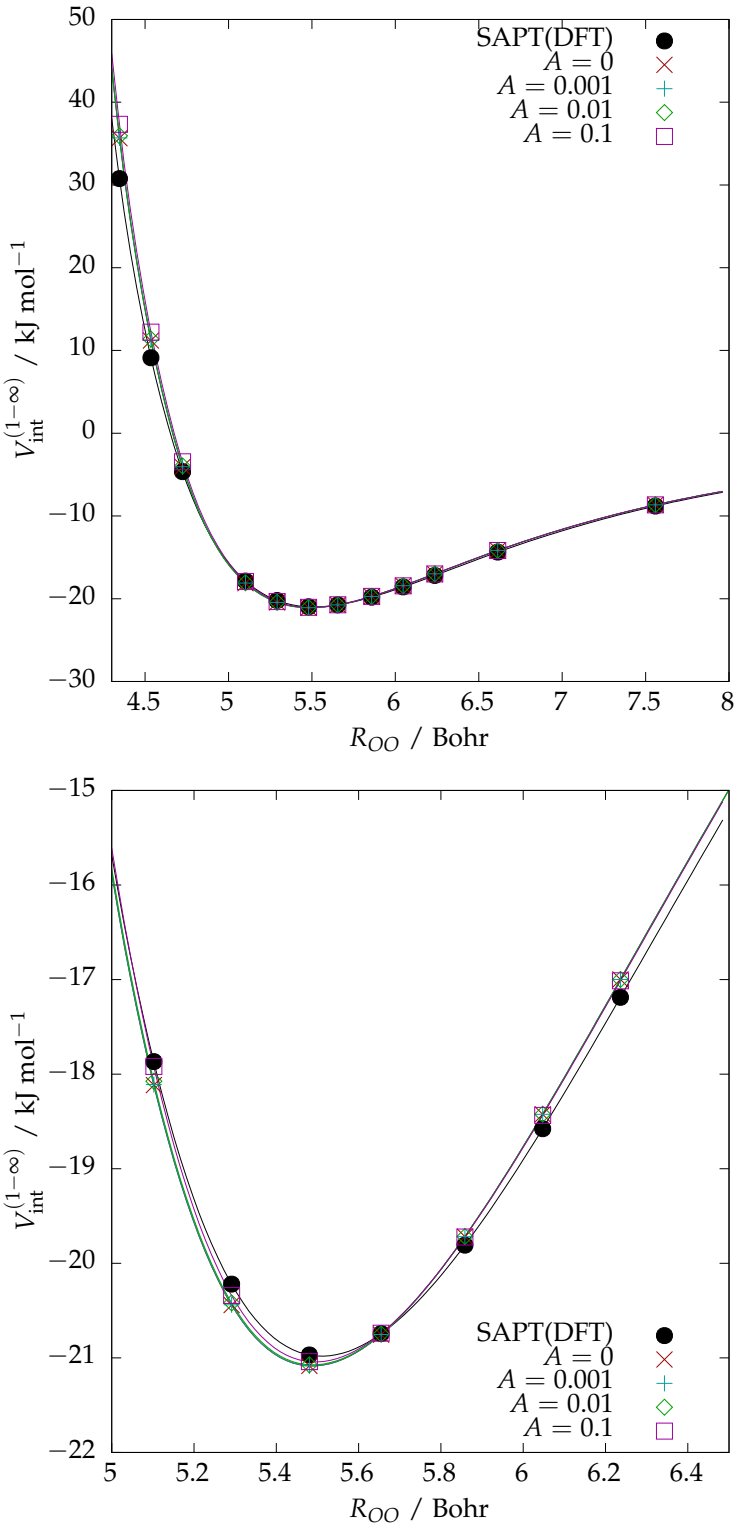



Figure 2.18: Profile plot of total interaction energy along the profile of the global minimum for different anchor strengths $A$ of $\alpha$'s and $\rho$'s, using weighting parameter $E_0 = 15$kJ mol$^{-1}$.

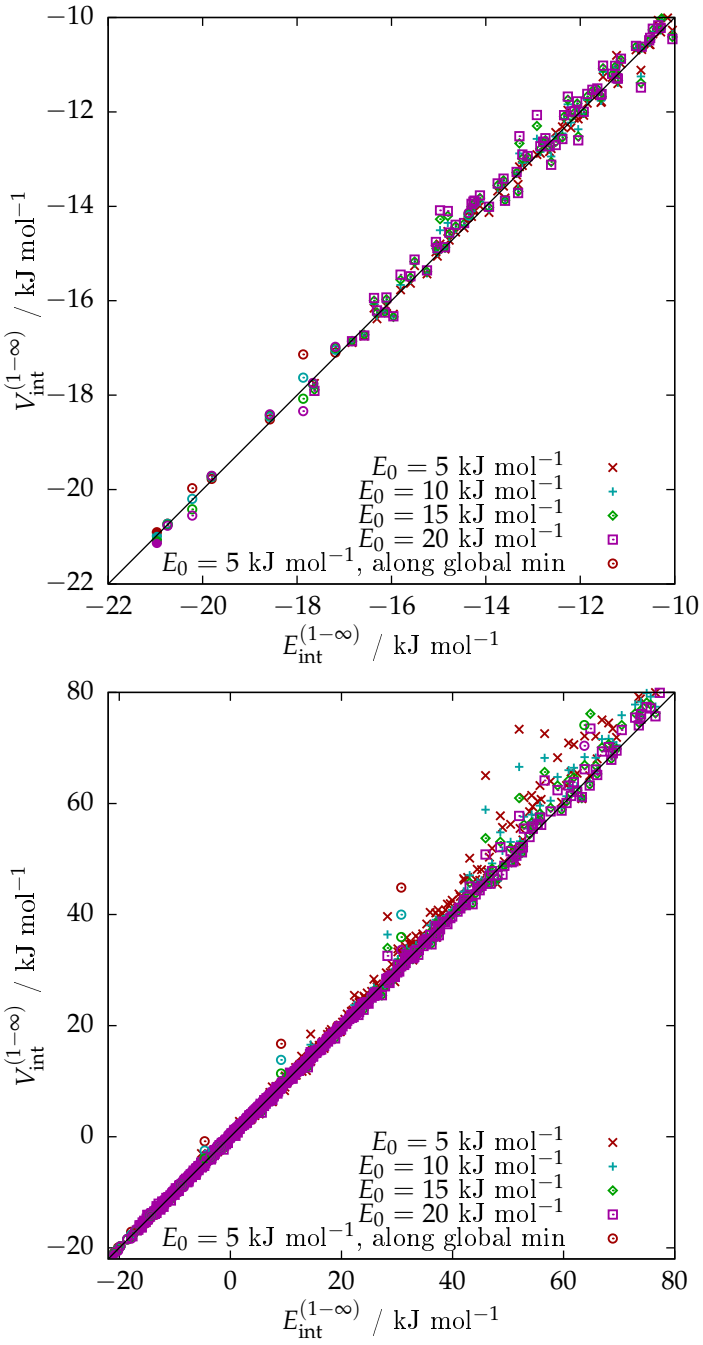



Figure 2.19: Total interaction energies varying the weighting parameter $E_0$ when fitting the water potential to the infinite-order energy, using anchors strengths of 0.01. Plots are over different energy ranges Circular markers indicate points along the profile of the global minimum.

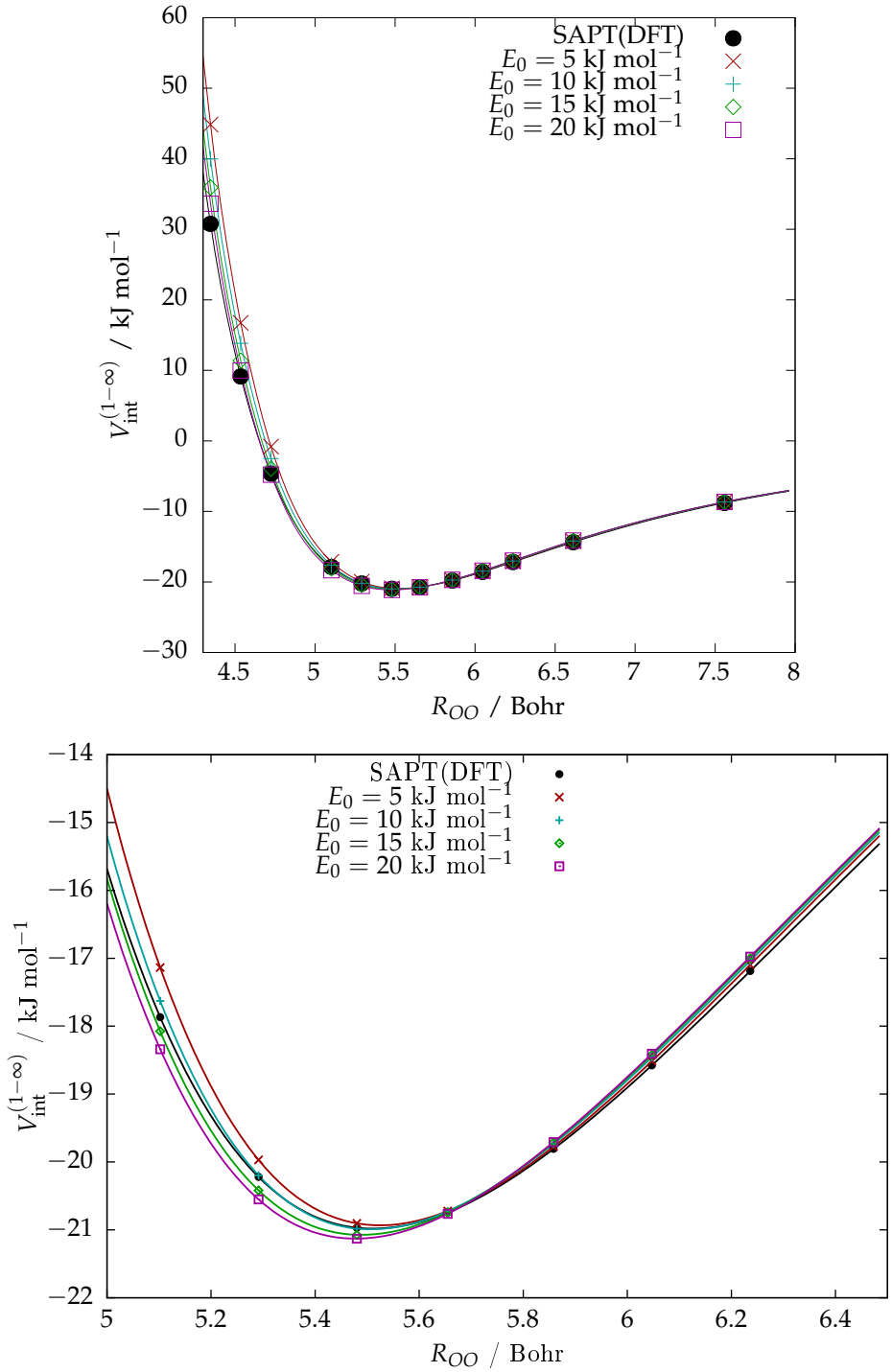



Figure 2.20: Profile plot of total interaction energy along the profile of the global minimum water dimer for different weighting parameters $E_0$, using anchor strengths of 0.01.

## 2.6 Fitting Short-range Energy for Methane

Fitting the short-range pair potential for methane initially used 2048 dimers running either side of the global minimum as described in Section 2.3.1. This set made it difficult to fit the repulsive wall, so an additional 600 points were used, having $d_{R1} = 1.8$ Bohr and $d_{R2} = 0$ Bohr.

### 2.6.1 Fitting the Short-Range Energy to First Order

The first-order fit was carried out without the use of any anchors, leaving only the Boltzmann weighting parameter $E_0$ to be determined. Geometries were categorised according to the total interaction energy from SAPT(DFT) calculations similar to Sec. 2.5, these categories are described in Table 2.13. At the global minimum the energy $E^{(1)}_{\text{exch}} + E^{(1)}_{\text{pen}}$ is close to 5 kJ mol$^{-1}$ so $E_0$ might be assumed to be somewhere close to this value. It was found that rmse was lower for larger values of $E_0$ and $E_0 = 10$ kJ mol$^{-1}$ was chosen; see Table 2.14 and Fig 2.21. The resulting parameters are given in Table 2.15. The terms $\rho^C_{10}$ and $\rho^C_{20}$ were included erroneously, however the fit assigned them very small values and they were simply removed from the next stage of the fit.

| $w = e^{-E/10\text{kJ Mol}^{-1}}$ | $E_{max}$ / kJ mol$^{-1}$ | Number of points $E_{int}$ within interval |
|---|---|---|
| $> 1$ | 0 | 1217 |
| $> 0.4, < 1$ | 9.163 | 877 |
| $> 0.1 < 0.4$ | 23.026 | 390 |
| $< 0.1$ | 73.067 | 168 |

Table 2.13: Method of splitting the reference points according to SAPT(DFT) total interaction energy, for the sake of fitting the short-range energy at first- and infinite-order. The lowest energy in the dataset was $-1.796$ kJ mol$^{-1}$.

| $E_0$ /kJ mol$^{-1}$ | 2 | 5 | 10 |
|---|---|---|---|
| $w > 1$ | 0.035 | 0.040 | 0.051 |
| $> 0.4, < 1$ | 0.416 | 0.277 | 0.243 |
| $> 0.1 < 0.4$ | 0.508 | 0.352 | 0.313 |
| $< 0.1$ | 0.605 | 0.428 | 0.380 |

Table 2.14: rmse in kJ mol$^{-1}$ from varying the weighting parameter $E_0$ when fitting the methane potential to the first-order energy. $w$ as in Table 2.13. $E_0 = 10$ kJ mol$^{-1}$ was chosen. Marker shapes indicate $w$, marker colours indicate the model.

### 2.6.2 Fitting the Short-Range Energy to Infinite Order

When fitting to infinite order, the $\rho_{10,00}$ and $\rho_{20,00}$ carbon terms from Table 2.15 were omitted and the remaining terms were used as anchor values with the weighting parameter starting at $E_0 = 2$ kJ mol$^{-1}$. For low energies there was very little difference in rmse as anchor strength $A$ was varied. There was little change in rmse or in the global minimum profile as $A$ was varied, see Table 2.16 and Figures 2.22 and 2.23, however it was noted that the output values of $\alpha$ had changed significantly from the input parameters for $A = 0.1$ and changed again at $A = 0.01$

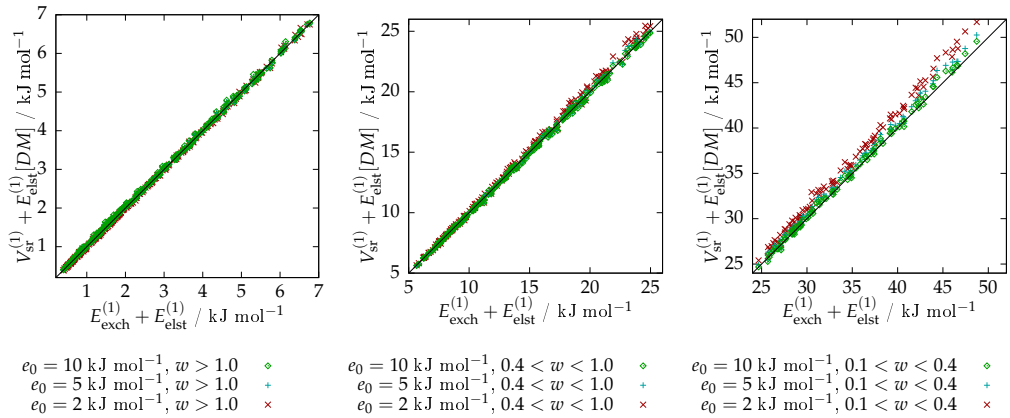



Figure 2.21: Scatter plot of the fitted first-order energy for the methane potential against $E_{exch}^{(1)} + E_{pen}^{(1)}$. $E_0 = 10$ kJ mol$^{-1}$ was chosen. Plots are for different values of $w$, i.e. different energy ranges, see Table 2.13 for details.

| Pair | $\alpha_{ab}$ Index (for $\rho_{ab}$) | $\rho_{ab}$ | |
|---|---|---|---|
| C-C | 2.146 | 00,00 | 5.648 |
| | | 10,- | 0.005 |
| | | 20,- | $-0.000$ |
| C-H | 1.867 | 00,00 | 4.228 |
| | | 10,- | 0.003 |
| | | -,10 | $-0.243$ |
| | | 20,- | $-0.000$ |
| H-H | 1.681 | 00,00 | 4.813 |
| | | 10,- | $-0.243$ |

Table 2.15: Potential parameters in a.u. for the methane potential fitted to $E_{exch}^{(1)} + E_{pen}^{(1)}$ with no anchors and isotropic C-H terms fixed to match C-C and H-H terms.

with little difference between output potentials for $A = 0.01$ and $A = 0.001$, hence $A = 0.01$ was chosen and not $A = 0.1$; see Table 2.17.

The weighting parameter $E_0$ was then fitted with $A = 0.01$. Although increasing the value of $E_0$ to 4 kJ mol$^{-1}$ would have resulted in a lower rmse (see Table 2.18), this would also lead to a poorer match along the profile of the global minimum and so $E_0$ was kept at 2 kJ mol$^{-1}$.

The terms in the $C_6$ dispersion model were then allowed to vary. As the dispersion model was a simple fit with Orient to SAPT(DFT) dispersion energies and not fitted with CamCASP using local polarisabilities as described in Sec. 2.2.2, the dispersion parameters are awarded less regard than was the case in the water-water or water-methane potentials. The dispersion parameter was subsequently given an anchor strength of $A_C = 0.1$, as this resulted in lowered rmse at higher energies (see Table 2.19 and Fig 2.27; with little effect on energies close to the global minimum (see Fig 2.26).

The final potential parameters for the repulsive wall and the dispersion model are given in Table 2.20.

| $A$ | 0.1 | 0.01 | 0.001 |
|---|---|---|---|
| $w > 1$ | 0.041 | 0.040 | 0.040 |
| $> 0.4, < 1$ | 0.156 | 0.151 | 0.150 |
| $> 0.1 < 0.4$ | 0.774 | 0.948 | 0.761 |
| $< 0.1$ | 9.964 | 10.302 | 7.805 |

Table 2.16: rmse in kJ mol$^{-1}$ from varying the anchor strength $A$ when fitting the the methane potential short-range energy to infinite order keeping $E_0 = 2$kJ mol$^{-1}$. $w$ as in Table 2.13, anchor parameters taken from Table 2.15.

| $A$ | (first order fit) | 0.1 | 0.01 | 0.001 |
|---|---|---|---|---|
| $\alpha_{CC}$ | 2.146 | 4.394 | 3.982 | 3.880 |
| $\alpha_{CH}$ | 1.867 | 2.110 | 5.260 | 4.380 |
| $\alpha_{HH}$ | 1.681 | 1.644 | 1.657 | 1.657 |

Table 2.17: Values of exchange-repulsion parameter $\alpha$ in a.u. from varying the anchor strength $A$ when fitting the short-range energy to infinite order keeping $E_0 = 2$kJ mol$^{-1}$.

| $E_0$ / kJ mol$^{-1}$ | 1 | 2 | 4 | 10 |
|---|---|---|---|---|
| $w > 1$ | 0.039 | 0.040 | 0.043 | 0.048 |
| $> 0.4, < 1$ | 0.164 | 0.151 | 0.145 | 0.143 |
| $> 0.1 < 0.4$ | 1.039 | 0.948 | 0.388 | 0.324 |
| $< 0.1$ | 13.320 | 10.302 | 3.977 | 3.403 |

Table 2.18: rmse in kJ mol$^{-1}$ from varying the weighting parameter $E_0$ when fitting the short-range energy to infinite order keeping $A = 0.01$. $w$ as in Table 2.13, anchor parameters taken from Table 2.15.

| $A_C$ | (fixed $C_6$) | 0.1 | 0.01 | 0.001 |
|---|---|---|---|---|
| $w > 1$ | 0.040 | 0.046 | 0.046 | 0.046 |
| $> 0.4, < 1$ | 0.151 | 0.146 | 0.152 | 0.152 |
| $> 0.1 < 0.4$ | 0.948 | 0.367 | 0.398 | 0.401 |
| $< 0.1$ | 10.302 | 3.412 | 3.427 | 3.424 |

Table 2.19: rmse in kJ mol$^{-1}$ from varying $A_C$, the anchor strength of dispersion coefficient. These fits used weighting parameter $E_0 = 2$ kJ mol$^{-1}$, anchor strength $A = 0.01$ and anchor values as shown in Table 2.15 for the exchange-repulsion parameters, and $C_6$ parameters as shown in Table 2.3. $w$ as in Table 2.13.

| Pair | $\alpha_{ab}$ | Index (for $\rho_{ab}$) | $\rho_{ab}$ | $C_{6,ab}$ |
|---|---|---|---|---|
| C-C | 2.179 | 00,00 | 5.553 | 22.079 |
| C-H | 1.899 | 00,00 | 4.187 | 10.095 |
| | | -,10 | $-0.204$ | |
| H-H | 1.648 | 00,00 | 4.782 | 4.872 |
| | | 10,- | $-0.276$ | |

Table 2.20: Final potential parameters for $V_{\text{sr}}^{(1)}$ and $V_{\text{disp}}^{(2)}$ in the methane-methane fit, all values in a.u.

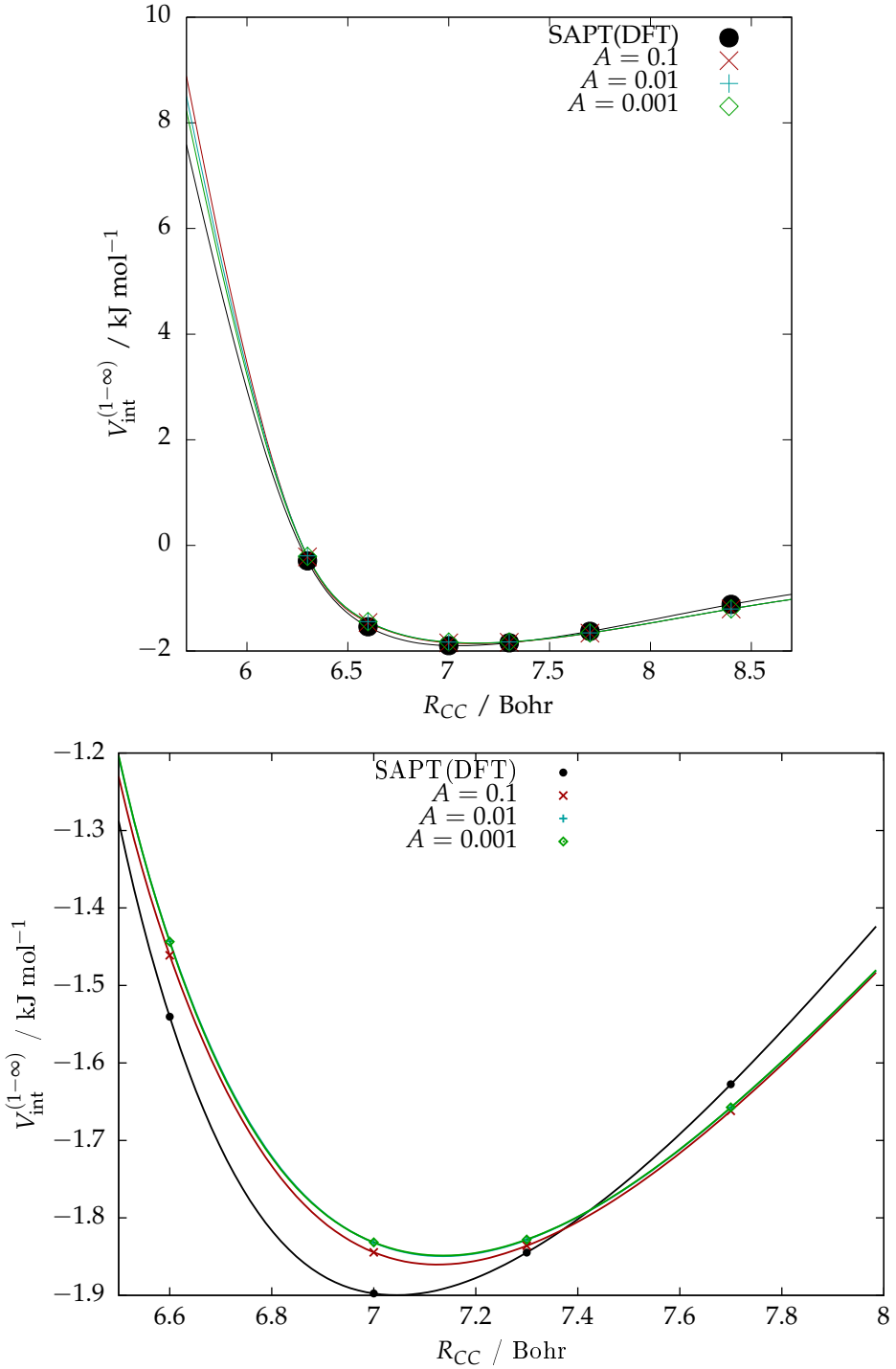



Figure 2.22: Profile plot of total interaction energy along the profile of the global minimum for the methane dimer for different anchor strengths $A$ of $\alpha$'s and $\rho$'s, using weighting parameter $E_0 = 2$kJ mol$^{-1}$ and anchor values as shown in Table 2.15.

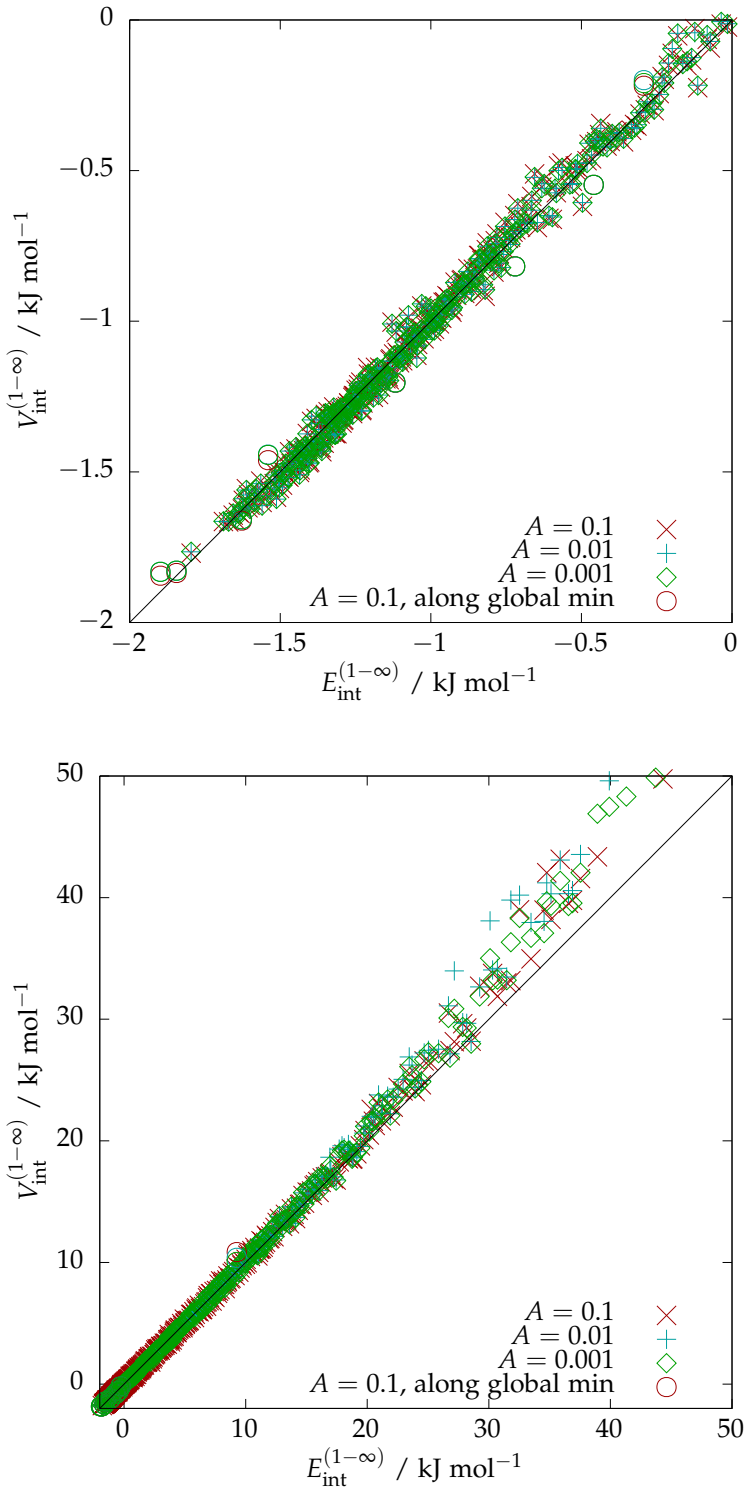



Figure 2.23: Total interaction energies varying the anchor strength $A$ of $\alpha$'s and $\rho$'s when fitting the methane potential to the infinite-order energy, using weighting parameter $E_0 = 2\text{kJ mol}^{-1}$ and anchor values as shown in Table 2.15. Plots are over different energy ranges. Circular markers indicate points along the profile of the global minimum.

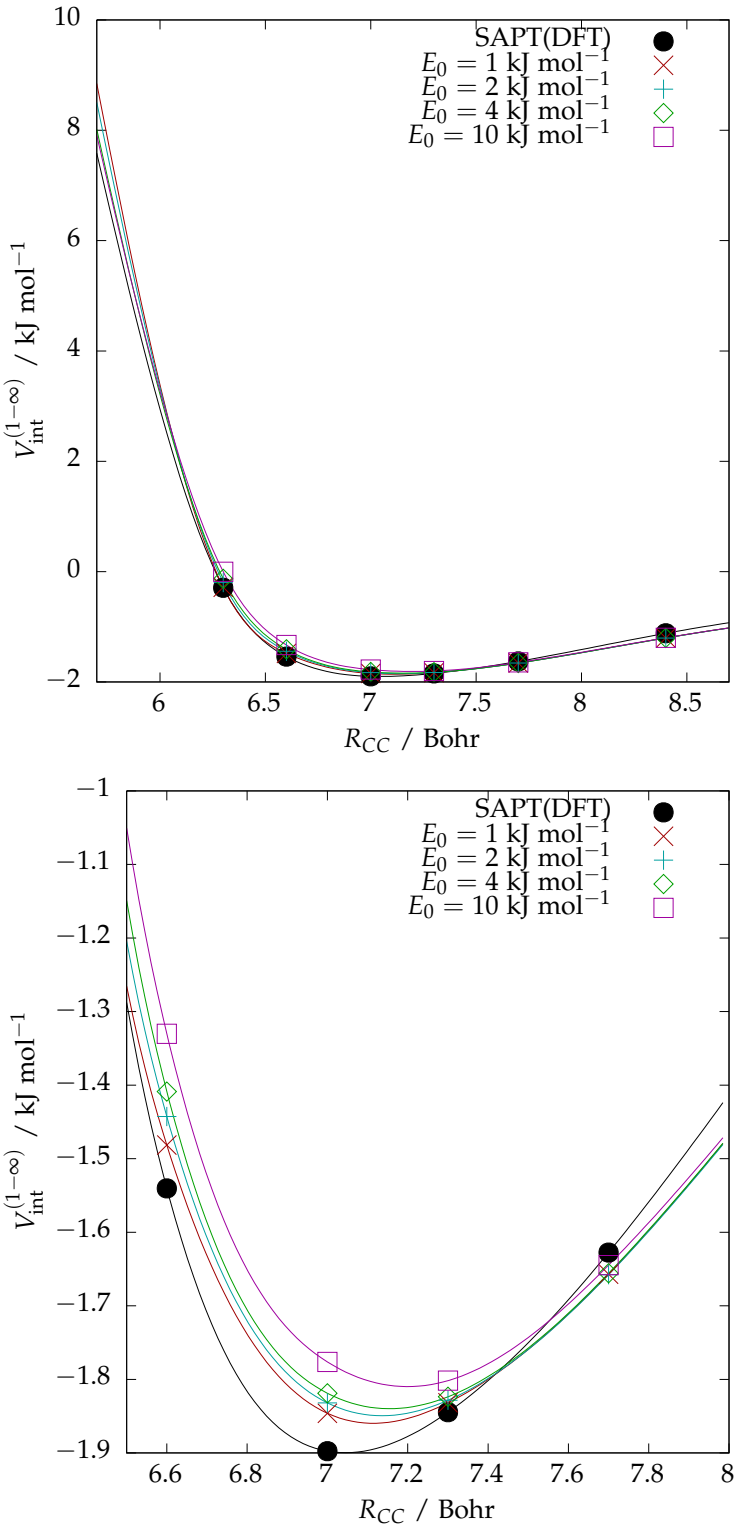



Figure 2.24: Profile plot of total interaction energy along the profile of the global minimum methane dimer varying weighting parameter $E_0 = 2$kJ mol$^{-1}$, using anchor strength $A = 0.01$ and anchor values as shown in Table 2.15. $E_0 = 2$ kJ mol$^{-1}$ was chosen.

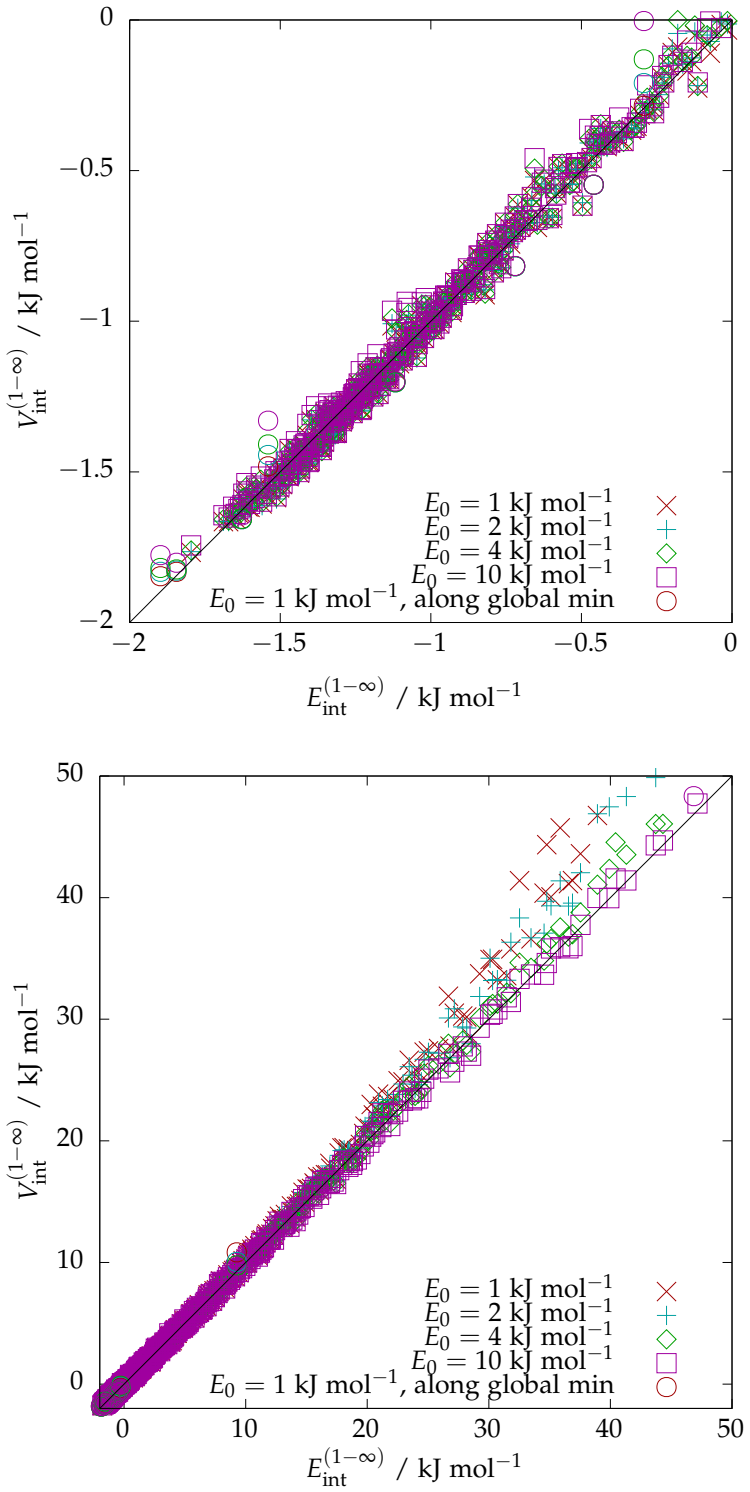



Figure 2.25: Total interaction energies varying the weighting parameter $E_0$ when fitting the methane potential to the infinite-order energy, using anchor strength $A = 0.01$ and anchor values as shown in Table 2.15. Plots are over different energy ranges. Circular markers indicate points along the profile of the global minimum. $E_0 = 2$ kJ mol$^{-1}$ was chosen.

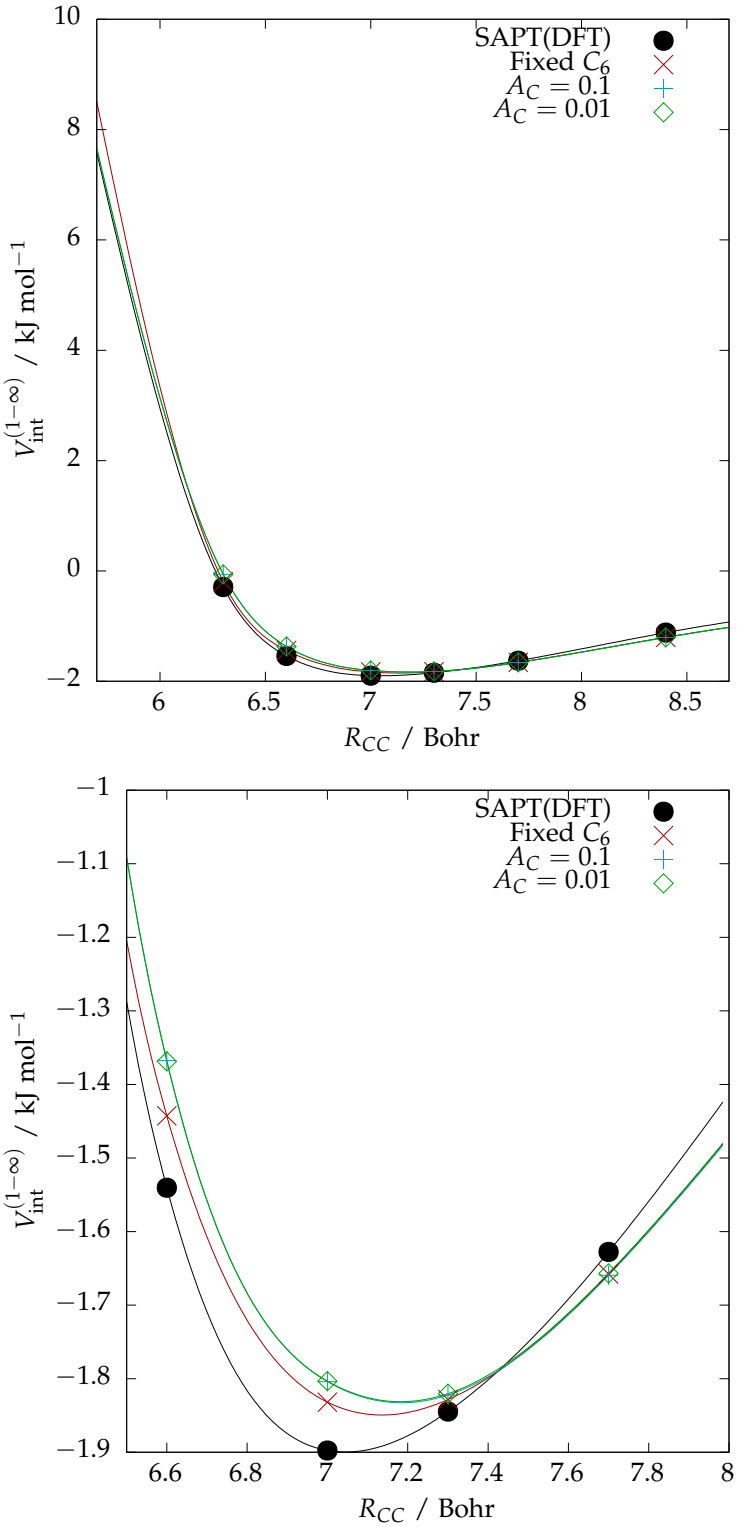



Figure 2.26: Profile plot of total interaction energy along the profile of the global minimum methane dimer varying $A_C$, the anchor strength of dispersion coefficient. These fits used weighting parameter $E_0 = 2$ kJ mol$^{-1}$, anchor strength $A = 0.01$ and anchor values as shown in Table 2.15 for the exchange-repulsion parameters, and $C_6$ parameters as shown in Table 2.3. $A_C = 0.1$ was chosen.

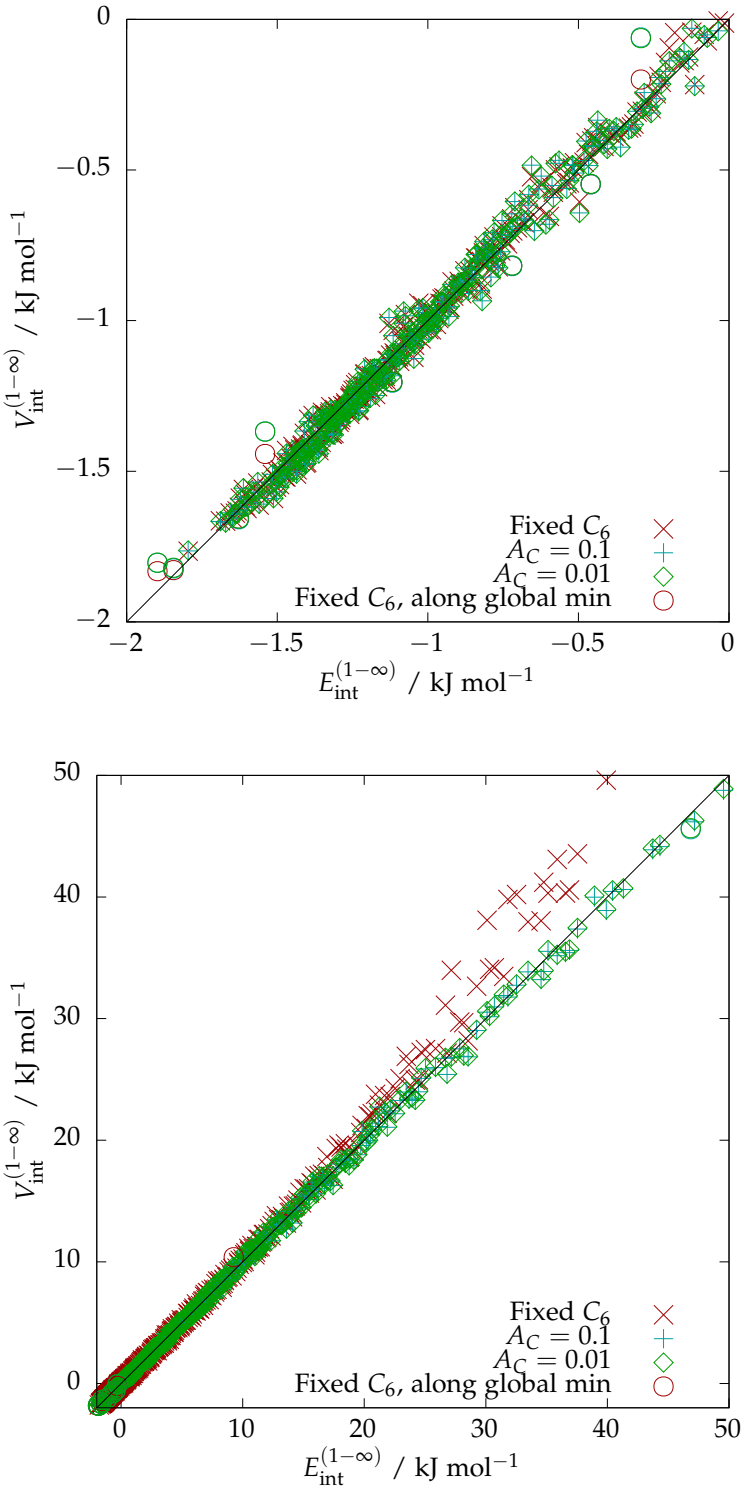



Figure 2.27: Total interaction energies for the methane potential varying $A_C$, the anchor strength of dispersion coefficient. These fits used weighting parameter $E_0 = 2$ kJ mol$^{-1}$, anchor strength $A = 0.01$ and anchor values as shown in Table 2.15 for the exchange-repulsion parameters, and $C_6$ parameters as shown in Table 2.3. Plots are over different energy ranges. Circular markers indicate points along the profile of the global minimum. $A_C = 0.1$ was chosen.

## 2.7 Fitting the Short-Range Energy for Methane-Water

The methane-water potential uses the same multipole and polarisation models as the water and methane potentials. The set of geometries used were made according to the same specifications as the methane potential; i.e. 2648 points as described in Sec. 2.6. As with the water pair potential, a satisfactory fit could not be attained until the $\rho^{O}_{22c}$ and $\rho^{H}_{11c}$ terms were included.

As with the methane-methane and water-water pair potentials the configurations used were categorised according to total energy, as shown in Table 2.21.

| $w = e^{-E/20\text{kJ Mol}^{-1}}$ | $E_{max}$ / kJ mol$^{-1}$ | Number of points $E_{int}$ within interval |
|---|---|---|
| $> 1$ | 0 | 1490 |
| $> 0.4, < 1$ | 9.163 | 977 |
| $> 0.1 < 0.4$ | 23.026 | 159 |
| $< 0.1$ | 73.067 | 19 |

Table 2.21: Method of splitting the reference points according to SAPT(DFT) total interaction energy, for the sake of fitting the short-range energy at first- and infinite-order. The lowest energy in the dataset was $-1.796$ kJ mol$^{-1}$.

### 2.7.1 Fitting the short-range energy to first-order

As an initial guess to the parameters, the shape function values from the water and methane pair potentials were used. (In the case of the $\rho_{00,00}$ terms Orient uses, these represent the sum of $\rho^{a}_{00}$ and $\rho^{b}_{00}$ terms so the average value can be taken.) At the time this was carried out a slightly different choice of parameter values for the short-range potential for water was preferred, and so the parameters used differ slightly from those shown 2.12. Whilst it would be preferable to start with the matching values, the difference is not likely to have any serious affect on the final potential, particularly as no anchors were used.

For $\alpha$ terms, the values were taken from an earlier fit to the overlap model. The starting parameters input into Orient are then as in Table 2.22. Within this section it is necessary to distinguish between the hydrogen atoms which appear in methane and the hydrogen atoms which appear in water; and so the former will continue to be denoted by the letter "H" and the latter will now be denoted by the letters "Hw".

The parameters were fitted with no anchors, meaning only the weighting parameter $E_0$ had to be decided. There as very little to choose between $E_0 = 15$ kJ mol$^{-1}$ and $E_0 = 25$ kJ mol$^{-1}$; 15 kJ mol$^{-1}$ was chosen as the rmse is slightly lower for negative energies (0.159 kJ mol$^{-1}$compared with 0.179 kJ mol$^{-1}$). The resulting potential is given in Table 2.24.

### 2.7.2 Fitting the short-range energy to infinite-order

Using the potential given in Table 2.24 for a set of anchor values, the potential was fitted to the full energy first varying the anchor strengths $A$ then the weighting parameter $E_0$. While varying the anchor strengths, $E_0$ was set to 5 kJ mol$^{-1}$, close to the value of $E^{(1-\infty)}_{\text{int}} - V^{(2-\infty)}_{\text{pol}}$ at the global minimum configuration. It was found that even for weak anchors the output potentials varied very little from the input potential: Using the weakest anchors ($A = 0.001$) the only terms which varied from the input potential by more than 0.1 were $\alpha^{CO}$ (from 2.207

| Site (pair) | $l_a$,$l_b$ | $\rho$ | $\alpha$ |
|---|---|---|---|
| O,C | 00,00 | 5.7758605 | 1.92296 |
| | 10,- | 0.015407 | |
| | 20,- | 0.017781 | |
| | 22c,- | $-0.136232$ | |
| O,H | 00,00 | 5.258275 | 1.878702 |
| | 10,- | 0.015407 | |
| | -, 10 | $-0.242856$ | |
| | -, 11c | 0.033224 | |
| | 20,- | 0.017781 | |
| | 22c,- | $-0.136232$ | |
| Hw,C | 00,00 | 4.6684945 | 1.932032 |
| | 10, - | $-0.119080$ | |
| | 11c,- | 0.031511 | |
| Hw,H | 00,00 | 4.250909 | 1.932032 |
| | 10, - | $-0.119080$ | |
| | -, 10 | $-0.242856$ | |
| | 11c,- | 0.031511 | |

Table 2.22: Starting parameters in the methane-water repulsive wall input into Orient (all values in a.u.).

| $E_0$ / kJ mol$^{-1}$ | 5 | 15 | 25 |
|---|---|---|---|
| $w > 1$ | 0.120 | 0.159 | 0.179 |
| $> 0.4, < 1$ | 0.693 | 0.445 | 0.428 |
| $> 0.1 < 0.4$ | 3.659 | 1.831 | 1.476 |
| $< 0.1$ | 14.966 | 6.368 | 5.087 |

Table 2.23: rmse in kJ mol$^{-1}$ from varying the weighting parameter $E_0$ when fitting the short-range energy to first-order for the methane-water model. $w$ as in Table 2.21.

| Site (pair) | $l_a$,$l_b$ | $\rho$ | $\alpha$ |
|---|---|---|---|
| O,C | 00,00 | 5.247938 | 2.206657 |
| | 10,- | $-0.001010$ | |
| | 20,- | 0.015517 | |
| | 22c,- | $-0.169833$ | |
| O,H | 00,00 | 5.323569 | 1.695928 |
| | 10,- | $-0.001010$ | |
| | -, 10 | $-0.241583$ | |
| | -, 11c | 0.033224 | |
| | 20,- | 0.015517 | |
| | 22c,- | $-0.169833$ | |
| Hw,C | 00,00 | 4.435371 | 1.688009 |
| | 10, - | $-0.095725$ | |
| | 11c,- | 0.038302 | |
| Hw,H | 00,00 | 4.153064 | 1.891707 |
| | 10, - | $-0.095725$ | |
| | -, 10 | $-0.241583$ | |
| | 11c,- | 0.038302 | |

Table 2.24: Repulsive wall parameters in a.u. from the fit to first-order for the methane-water model.

to 2.038), $\rho^{CHw}_{00,00}$ (4.435 to 4.727) and $\rho^{Hw}_{10}(H)$ (-0.096 to -0.260). Given these relatively small changes, and the very small differences in rmse (Table2.25), the weak anchors were kept.

With the anchor strengths decided, $E_0$ was increased to 15 kJ mol$^{-1}$. This resulted in a poorer fit at the minimum, see Fig 2.28, and so $E_0$ was kept at 5 kJ mol$^{-1}$. Finally, the first dispersion coefficient $C_6$ was allowed to vary for each pair by introducing an anchor of strength $A_C = 0.01$. This gave a better fit overall with little change to the global minimum (Fig 2.28); and little change to the $C_6$ coefficients themselves. This gave the final water-methane pair potential, with exchange-repulsion and $C_6$ dispersion coefficients as shown in Table 2.26 and other dispersion coefficients as shown previously in Table 2.5.

| $A$ | 0.1 | 0.01 | 0.001 |
|---|---|---|---|
| $w > 1$ | 0.187 | 0.184 | 0.183 |
| $> 0.4, < 1$ | 0.730 | 0.720 | 0.719 |
| $> 0.1 < 0.4$ | 1.748 | 1.741 | 1.770 |
| $< 0.1$ | 4.562 | 4.933 | 4.651 |

Table 2.25: rmse in kJ mol$^{-1}$ from varying the anchor strength $A$ when fitting the short-range energy to infinite order for the methane-water model keeping $E_0 = 5$kJ mol$^{-1}$. $w$ as in Table 2.21, anchor parameters taken from Table 2.24.

| Pair | $\alpha$ | $l_a$ | $l_b$ | $\rho$ | $C_6$ |
|---|---|---|---|---|---|
| O,C | 2.206657 | 00 | 00 | 5.375834 | 16.90574 |
| | | 10 | | −0.015507 | |
| | | 20 | | −0.064900 | |
| | | 22c | | −0.151909 | |
| O,H | 1.695928 | 00 | 00 | 5.265648 | 8.09713 |
| | | 10 | | −0.002486 | |
| | | | 10 | −0.331392 | |
| | | 20 | | 0.015588 | |
| | | 22c | | −0.144187 | |
| Hw,C | 1.688009 | 00 | 00 | 4.769332 | 4.77852 |
| | | 10 | | −0.100053 | |
| | | 11c | | 0.081363 | |
| Hw,H | 1.891707 | 00 | 00 | 4.109502 | 2.47039 |
| | | 10 | | −0.292882 | |
| | | | 10 | −0.238494 | |
| | | 11c | | 0.009255 | |

Table 2.26: Final parameters for the short-range energy and $C_6$ coefficients for the methane-water potential, all values in a.u.

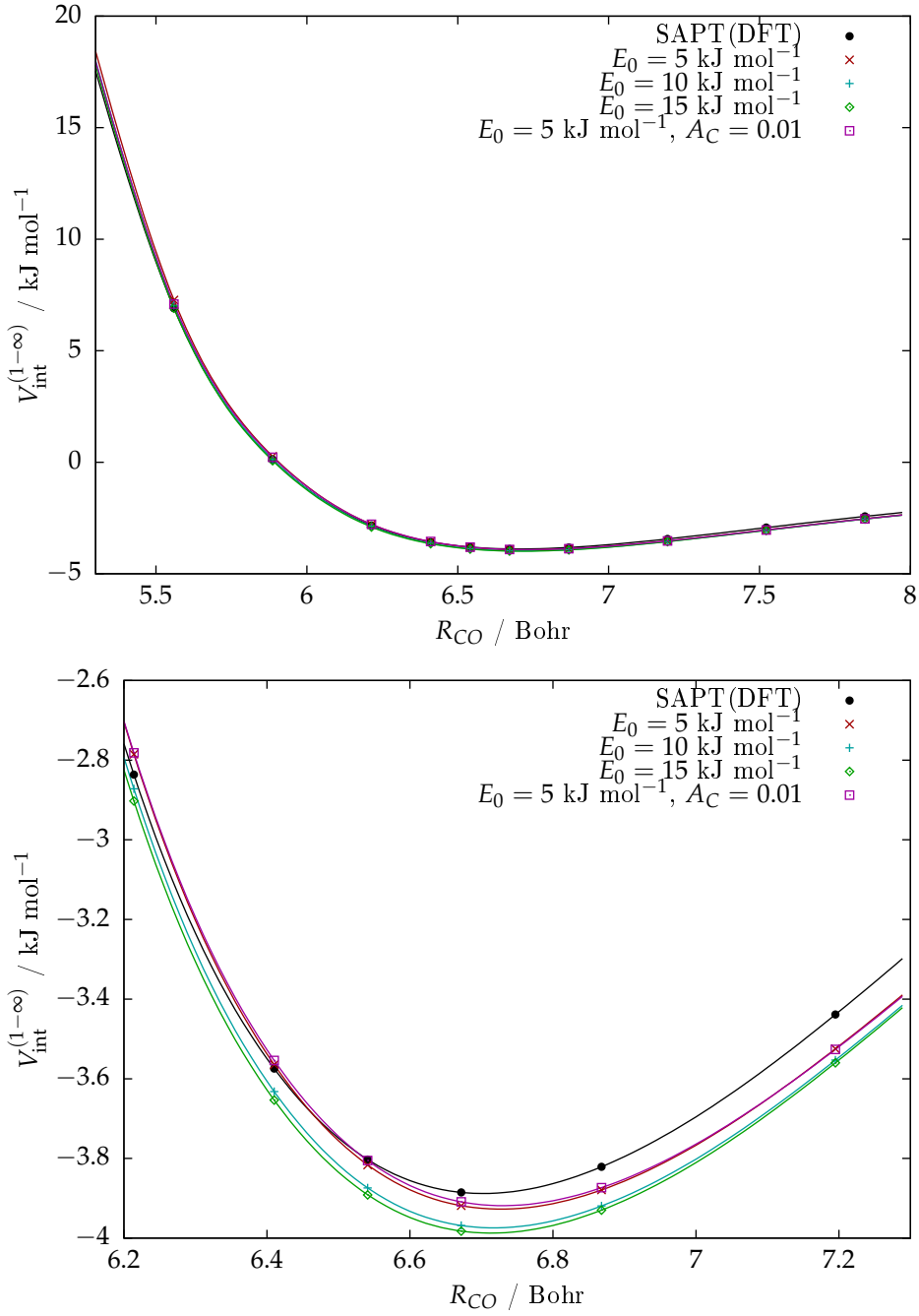



Figure 2.28: Profile plot of total interaction energy along the profile of the global minimum methane-water dimer varying weighting parameter $E_0 = 2$kJ mol$^{-1}$, using anchor strength $A = 0.001$ and anchor values as shown in Table 2.15. $E_0 = 5$ kJ mol$^{-1}$ was chosen and then $C_6$ anchor $A_C$ was set to 0.01.

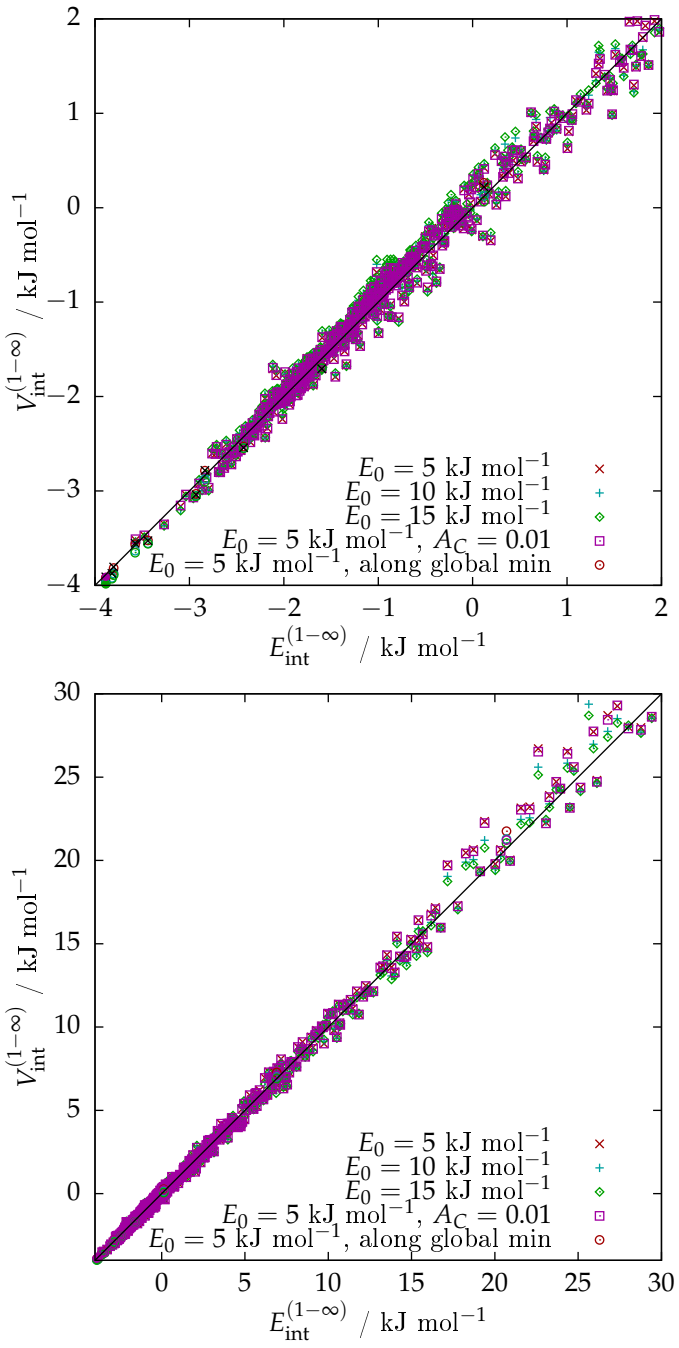



Figure 2.29: Total interaction energies varying the weighting parameter $E_0$ when fitting to the infinite-order energy for the methane-water model, using anchor strength $A = 0.001$ and anchor values as shown in Table 2.15. Plots are over different energy ranges. Circular markers indicate points along the profile of the global minimum. $E_0 = 5$ kJ mol$^{-1}$ was chosen and then $C_6$ anchor $A_C$ was set to 0.01.

## 2.8 Summary

In this section potentials are fitted to SAPT(DFT) reference energies for water, methane, and water-methane pairs. These models have distributed multipole models up to rank 4, with anisotropic distributed polarisabilities and isotropic dispersion models both up to rank 3 (except the methane-methane potential, which has only a $C_6$ term for dispersion), and use anisotropic terms to describe the exchange-repulsion.

When modelling water, it was found that an accurate reproduction of the global minimum without using it as a fit point was only possible when the repulsive wall was allowed anisotropy up to the 22c-level for oxygen and 11c-level for hydrogen. It also became clear that the DMA2 method of generating distributed multipoles was unreliable in the case of water and instead the DF-ISA method was used. Neither of these features appear in published water potentials currently in use and in the case of water and water-methane pair potentials there are no published models using a comparably extensive form.

The number of fit points used was quite small, 2048 for water and 2648 for methane and water-methane, with all geometries being randomly generated between given radii. In spite not using the global minimum directly as a data point in the fit, all potentials reproduced the depth of the global minimum to within 0.1 kJ $mol^{-1}$. Further evaluation of the models is given in Chapter 4.

# Chapter 3

# Simplification of Potentials

The anisotropic pair-potentials developed in the previous chapter had to be simplified for use in MD simulations with DL_POLY4. In brief this required replacing the repulsive wall shape functions with a single (isotropic) term for each atom pair; replacing the multipole models with point charges only; and replacing the damped rank 3 anisotropic polarisation model with an undamped rank 1 isotropic model, or with no polarisation at all. The isotropic dispersion model from the previous section was kept, and were incorporated into a tabulated potential with the repulsive wall. The simplified models used rigid bodies with the same monomer geometries as the previous chapter.

The most significant compromises related to the polarisation model for water. Polarisation is available in DL_POLY4 only via a charge-on-spring "shell model". While there is no reason in principle that a water molecule cannot be made with a shell model on each atom site this was not possible in practice and we would ultimately be limited to a single polarisable site. Polarisation was also affected at short distances in particular by the lack of damping and by the effect of including extra charge sites in the water model.

## 3.1 Description of Simplification Processes

### 3.1.1 Electrostatic energy

The multipole moments were replaced by point charges. In the case of methane, which is close to isotropic and has only small multipole moments (see Table B.5), the ISA charges were used. In the case of water, extra charge sites were introduced. The locations for extra sites were chosen by hand and the charges on all sites were then chosen using the program MULFIT [120]. MULFIT adjusts charge values so that those within a given radius of each site give a potential best matching the multipole moments over a given region about that site. For the water potential, the radius was large enough for all charge sites to be included; and the region used for reference was 2.27 to 5.67 Bohr from each charge site.

### 3.1.2 Exchange-repulsion energy

The exchange-repulsion energy is now modelled with only two parameters for each atom-pair, taking the form $Ge^{-\alpha_{ab}(R_{ab}-\rho_{ab})}$ where unlike in the previous chapter $\rho_{ab}$ takes a single numerical value. This fit is carried out using Orient similar to the previous section. It should

be expected to give results similar to the averaged values from the anisotropic potential, that is to say that $\alpha_{ab}$ should be similar and $\rho_{ab}$ should be similar to $\rho^a_{00} + \rho^b_{00}$, consequently these terms can be used as anchor parameters and the fit carried out to the infinite-order short-range energy in a single step. The only choice to be made by hand then is the relative weight assigned to each point. As previously mentioned, fitting to the infinite-order energy in Orient requires subtracting the polarisation model first, but each fitting point was still assigned a statistical weight as a function of its total interaction energy.

### 3.1.3 Polarisation energy via the shell model

The only way polarisation effects can be explicitly included in DL_POLY is via the "shell model", which uses charges to recreate an induced dipole. A positively charged core is encased by a negatively charged shell each of magnitude $q_s$. As the core is enclosed completely, it experiences no force due to the charge of the shell; however it is made to interact with the shell via an harmonic "spring" force $F_\alpha = kd_\alpha$ where $d$ is the separation between the core and the centre of the shell and $k$ is the spring constant.

If the charges are separated by some force due to an external field $E$ then

$$F = q_s E = kd \tag{3.1}$$

This charge separation also gives a dipole moment

$$\mu = q_s d = \alpha E \tag{3.2}$$

Combining these equations relates the polarisability to the shell charge and spring constant by

$$\alpha = q_s^2/k \tag{3.3}$$

This determines the choice of spring constant for a given shell charge (or vice versa).

For some site $a$ on molecule $A$ producing a field at site $b$ on molecule $B$, provided that the core-shell separation is always small compared with the distance between sites then the energy is given by a term similar to the term used for $V^{(2-\infty)}_{\text{pol}}$, with the exception that there is no damping function and that the polarisability $\alpha^b_{\alpha\beta}$ is now the isotropic term $\alpha^b = \frac{(q^b_s)^2}{k^b}$. Written in terms of chosen parameters $q^b_s$ and $k^b$, this is

$$V^{(2-\infty)}_{\text{pol}} = \frac{(q^b_s)^2}{2k^b} V^a_\alpha V^a_\alpha \tag{3.4}$$

When the atom already has some charge $q$ attributed to it, it is conventional to make the core charge $q + q_s$ and the shell charge $-q_s$ where $q_s$ is positive. This comes from the idea that the core can in some sense represent the shielded nuclear charge and the shell the valence charge: As intermolecular interactions are typically attributed to electrons, all non-coulomb interactions are conventionally taken to occur between shells and not cores. Although we adopt this convention here we discard the idea of the core and shell having physical meaning; acknowledging that we use it only as a means to replicate the point polarisabilities of a rigid model. Note that in order to imitate point polarisabilities, shells and cores from different atom sites in the same molecule also do not interact.

DL_POLY allows for two methods of using the shell model: The adiabatic shell model, in which the shell has some small mass; and the relaxed shell model, in which it doesn't. For the relaxed shell model the shell positions are adjusted at every timestep to minimimse the system energy. For this reason, the adiabatic shell model is faster in DL_POLY and is the method chosen.

#### 3.1.3.1 Existing water shell models

The shell model has been used to provide polarisability to water models in the past, generally using a core-shell unit for the oxygen site.

The first application appears to be de Leeuw and Parker (1998) [121], in a potential developed for simulating liquid water interacting with a magnesium oxide surface, with the intent that the interaction with other inorganic solids could also be studied. This was a flexible water model with a core-shell unit on the oxygen site, and is the only model listed here to use the adiabatic shell model. The model was fitted to the water dimer and was able to reproduce the radial distribution function and energy of vapourisation; the density of liquid water using this model was too large.

Van Maaren and van der Spoel (2001) [122] developed several water models which functioned essentially as shell models but had the shell charge connected to a dummy site positioned close to where the extra site "M" appears on TIP4P models. These models included both rigid and flexible models and also compared isotropic shell models to shell models with an anisotropic spring constant. Although the parameters were fitted to match the energy and density of liquid water, they also reproduced the dimer well. It was found that anisotropic polarisabilities did not have much advantage over isotropic polarisabilities and that, in this case, rigid models were preferable to flexible.

The SWM4-DP [123] and SWM4-NDP [124] models both use four sites similar to the TIP4P arrangement with a core-shell unit at the oxygen site, such that the oxygen has no net charge (the hydrogen sites are positive and the extra site M is negative). The two models differ in that SWM4-DP has a positive shell and SWM4-NDP has a negative shell. They highlight that this could be a significant difference since if a given intermolecular interaction causes the shell to move towards the site M in one model, the same interaction will cause the shell to move away from the site M in the other; but both cases accurately reproduce vaporization energy, density, static dielectric constant and self-diffusion coefficient.

Both SWM4-DP and SWM4-NDP have a polarisability lower than the experimental value of 1.44 $Å^3$ at 1.04252 $Å^3$ and 0.97825 $Å^3$ respectively. They suggest this compensates for Pauli exclusion effects. In our own case some reduction in polarisability was also required, see Sec. 3.2.2.

The SWM4-NDP model was extended into a six-site model SWM6 [125] by including two extra sites representing the oxygen lone pairs. The polarisability had to be slightly reduced again to 0.88 $Å^3$, likely due to the dipole moment due to the new sites. With the inclusion of the lone pairs, which were negative, the oxygen core-shell unit now had a positive total charge. This six-site model gave similar bulk properties to the SWM4-NDP model but gave better results for the dimer structure and for the energies of water clusters, though in the particular case of water hexamers it had worse energetic ordering.

#### 3.1.3.2 Damping in the shell model

It is possible in DL_POLY to include quartic terms, i.e. extend the shell-core interaction to $k_1 d^2 + k_2 d^4$, which serves to prevent excessively large dipole moments. However, this depends on total coulomb force at core-shell unit $a$, i.e. both the separation and magnitude of interacting charges. The shell model does not allow for damping as a function only of atom-atom separation $R_{ab}$, as in Tang-Toennies damping. Because of difficulties relating these two forms of damping, no damping was used.

### 3.1.4 Accounting for polarisation via the electrostatic model

For bulk simulations the effect of polarisation, i.e. a total lowering of energy, can be accounted for to some extent by altering the electrostatic model. In particular the dielectric constant of water is known to be $\epsilon = 78.4$ at ambient pressure and temperature [126]. The charge distribution given this dielectric can be found using DFT calculations, and this can be used to give a new set of multipoles using the ISA method. These can be used with Mulfit to give a new set of point charges which are designed to work with the previously determined exchange-repulsion and dispersion models.

This is only an approximate method and the value of $\epsilon$ changes depending on the temperature and the medium, for instance at 273 K the dielectric constant of ice Ih is 94 whereas for sI clathrate it is around 58 [45].

#### 3.1.4.1 Selection of correct electrostatic model

Suppose a multipole moment $Q_t^a$ of rank $t$ at some site $a$ in molecule $A$ interacts with an external field[1], so that the multipoles at site $a$ change by $\Delta Q_t^a$. Then, the energy due to interaction of the molecule with the field changes by (using the summation convention for sites and moments)

$$\Delta E_f = \Delta Q_t^a V_t^a \tag{3.5}$$

$$= -\alpha_{tt'}^{aa'} V_t^a V_{t'}^{a'} \tag{3.6}$$

for polarisability matrix $\alpha_{tt'}^{aa'}$.

There is also an energy penalty paid for deforming the charge distribution of $A$, which is given by

$$\Delta E_A = \frac{1}{2}\alpha_{tt'}^{aa'} V_\alpha^a V_\beta^a \tag{3.7}$$

The partial cancellation of $\Delta E_f$ by $\Delta E_A$ gives the familiar term for the polarisation energy.

When attempting to replicate the polarisation energy by increasing the multipole moments, there is no penalty paid and so the correct energy can be recovered only by introducing a change in moment of $\Delta Q_t^a/2$ instead of $\Delta Q_t^a$. This method was used early on by Berendsen et al. [127] when parameterising the SPC/E model and has also been used, for example, by Welch et al. to estimate the induction energy of organic crystals [128].

We consider though that when we change the multipole moments by a fixed amount, we ought to be altering the internal energy of molecule $A$ by a fixed penalty. In contrast, the

[1] In this notation $Q_0^a = q^a$, $Q_2^a = \frac{1}{3}\Theta_{\alpha\beta}^a$, etc.

change in interaction energy is dependent on the position of site $a$. By changing the multipole moments by $\Delta Q_t^a/2$, then, we are making the penalty also depend on the position of site $a$.

If the value for $\Delta Q_t^a$ is found in a low-energy configuration we expect the polarisation energy, and consequently the penalty, to be relatively large. By altering (in general reducing) the penalty but not the multipole moments as the geometry changes, we might expect the low-energy configuration to become less distinct.

With this argument in mind, it was not clear to us whether the charge model should include the full term $q^a + \Delta q^a$ or only $q^a + \Delta q^a/2$. Both models were tested in simulation and in clusters, see Chapter 4.

There is some ambiguity over the method of determining $q^a + \Delta q^a/2$: Either the charges could be fitted to multipole moments first and then the midpoint of $q^a$ and $q^a + \Delta q^a$ found, or the midpoint of the full multipole moments $Q_t^a$ and $Q_t^a + \Delta Q_t^a$ found and then used to fit charges. The latter approach was taken, fitting to the SAPT5-s charge arrangement as discussed in Sec 3.2, and the resulting charge models are given in Table 3.1.

| | ISA($\epsilon = 1$) | (ISA($\epsilon = 1$)+ISA($\epsilon = 78.4$))/2 | ISA($\epsilon = 78.4$) |
|---|---|---|---|
| O | 0.11923 | −0.00884 | −0.13461 |
| H | 0.52245 | 0.56071 | 0.59852 |
| L | −0.58207 | −0.55599 | −0.53121 |

Table 3.1: Charge values in a.u. for models made with Mulfit using the SAPT-5s charge arrangement, with the input moments coming from ISA L4 models for $\epsilon = 1$ (i.e. a vacuum), $\epsilon = 78.4$ as with water at ambient conditions, and from the midpoint of these two ISA L4 models.

#### 3.1.4.2 Comparisons with 5 molecules in ice Ih in a dielectric

The ISA partition scheme also allows multipoles to be fitted to each atom for more than one molecule. This means that we do not necessarily have to assume that a single water molecule is placed in a dielectric medium resulting from all surrounding molecules, but can also surround a water molecule by its nearest neighbours first and then place the resulting cluster in a dielectric. Here the former is compared with the latter. Although there was not adequate time, the ISA method would also allow for multipoles to describe a water molecule surrounded by the next layer of nearest neighbours, and so on, presumably these would converge.

In water models it is common to try and account for the effects of polarisation at ambient temperature, for this reason a dielectric constant of $\epsilon = 78.4$ has been used, but this enhanced charge model also plays a particularly crucial role in the stability of frozen water, so the surrounding molecules are positioned as in ice Ih.

Ice Ih follows Bernal and Fowler's ice rules ensuring that one hydrogen lies between each pair of oxygen atoms. In practice though the hydrogen positions are also disordered. To account for this, the oxygen sites are put in place first with the hydrogen atoms oriented randomly, and their positions adjusted via a Monte Carlo procedure until the ice rules are obeyed.

Four such arrangements were chosen arbitrarily and are shown in Fig 3.1. The fitted ISA multipoles gave a total charge of less than 0.01 a.u. for all arrangements except for the one shown in Fig 3.1(b), which had a charge of around 0.07 a.u., total charges are given in Table 3.3. Another concern was that the resulting multipole models were not symmetric, but as the

difference in charges between hydrogen atoms was always less than 0.01 a.u. this was not deemed to be significant.

To compare multipole models an isodensity surface is used. The grid used consisted of 6408 triangles at an electron density of $10^{-2}$ a.u. giving the electrostatic energy for an interaction between the multipole model and a unit charge at 3206 points. The difference in electrostatic energy between the multipole models for a single molecule in a dielectric and a single molecule surrounded by four other molecules in a dielectric is given tabulated in Table 3.3 and visualised in Fig 3.2. The site charges are also given in Table 3.2.

Plots were made along the profile of the global minimum using the water potential developed in the previous chapter but replacing the multipole model with the model for either a single molecule in dielectric or the central molecule in one of these clusters in dielectric. These are shown in Fig 3.3 and show that the single molecule in dielectric gives significantly lower energies. A scatter plot of $E_{\text{int}}^{(1-\infty)}$ using these models also shows that the single molecule gives lower energies for most, but not all points, see Fig 3.3. Plots for the set of 2048 reference energies used in the short range fit are given in Fig 3.4.

| Atom | ISA($\epsilon = 78.4$) | A | B | C | D |
|---|---|---|---|---|---|
| O | $-0.9757$ | $-0.9944$ | $-0.9463$ | $-0.9837$ | $-0.9856$ |
| H(average) | 0.48878 | 0.49859 | 0.50779 | 0.4955 | 0.49443 |

Table 3.2: Atomic charges in a.u. using the ISA multipole model and water in a dielectric; compared with water surrounded by four other water molecules in a dielectric as shown in 3.1.

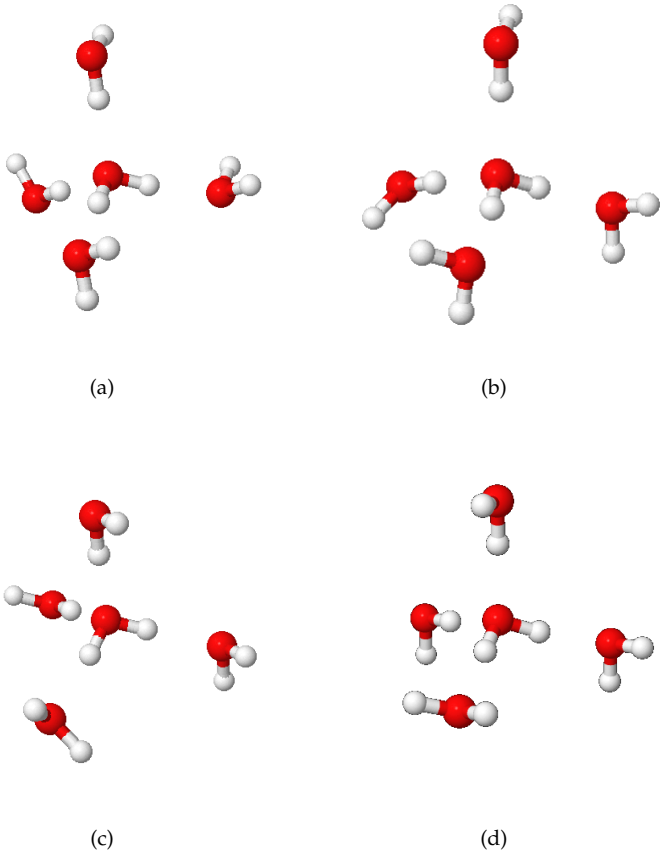
| | A | B | C | D |
|---|---|---|---|---|
| Total Charge | 0.002779 | 0.069222 | 0.00729 | 0.003263 |
| Max. Dif. | 42.369 | $-8.037$ | 35.101 | 44.108 |
| Min. Dif. | $-54.259$ | $-112.309$ | $-63.786$ | $-59.436$ |
| rmse | 19.564 | 59.046 | 21.289 | 19.857 |

Table 3.3: Differences in kJ mol$^{-1}$ between the rank 4 multipole model from a single water molecule in a dielectric medium and four arrangements of a water molecule surrounded by two donor and two acceptor molecules in a dielectric medium, for a surface electron density of 0.01. Total charges in a.u. for each set of multipoles are also given. Arrangements are shown in Fig 3.1. Positive difference indicates more positive charge for single molecule in dielectric.

## 3.2 Simplifying the Water Potential

In simplifying the water potential several compromises had to be made to account for the lack of polarisation damping and avoid the "polarisation catastrophe". With a view to find the best match to the multipole model the first attempt to fit the potential used a seven-site charge model. Unfortunately while the charges did well to match the electrostatic potential and the multipole moments at each atom site; they themselves produced multipoles which greatly increased the polarisation energy. To reduce this effect a five-charge model had to adopted; and the polarisabilities had to be slightly reduced from the dipole polarisabilities given in the anisotropic rank 3 model.

It was assumed that when transferring the potential to DL_POLY that undamped isotropic point polarisabilities on each atom site could be adequately reproduced with the shell model,



(a) (b)

(c) (d)

Figure 3.1: Arrangements of 5 water molecules taken from ice Ih structure made using the ice rules and a Monte Carlo procedure. When ISA multipoles were fitted to these four arrangements placed in a dielectric medium, all gave positive total charges below 0.01 a.u., except for Fig 3.1(b), which had a charge of around 0.07 a.u. The multipoles fitted to the central molecule were compared with the multipoles of a single molecule in the same dielectric, see Table 3.3 and Fig 3.2.

and the simplification process was conducted with this in mind. Ultimately, only a single polarisable site was possible, placed at the oxygen atom. Furthermore, the conversion from point-polarisable site to shell model was not an exact one. This is discussed further in Sec 5.2.

### 3.2.1 Charge models

The charge arrangements used in the water potential took three forms:

- A seven site model with one site on either side of the oxygen atom along the molecule's line of bisection, labelled types LVa and LVb, and a pair out of the plane at the tetrahedral angle as in, for example, the TIP5P model [129], labelled as type LT. See Fig 3.5.
- A five site model with only atom sites and tetrahedral LT sites.
- A five site model with charges positioned as in the SAPT-5s potential [14], see Fig 3.6.

In all cases charge values were found using MULFIT. The motivation for charges LT is to accommodate in some sense the oxygen lone pair and the preference for water to arrange

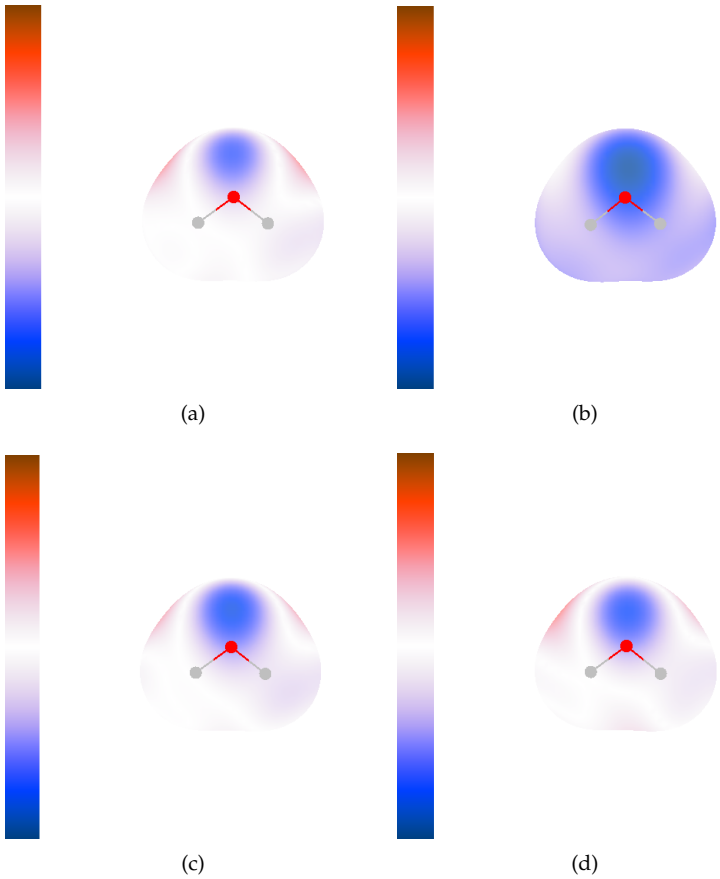



Figure 3.2: Differences in electrostatic energy between the rank 4 multipole model from a single water molecule in a dielectric medium and four arrangements of a water molecule surrounded by two donor and two acceptor molecules in a dielectric medium, interacting with a unit charge at a surface electron density of 0.01. Red indicates a positive difference i.e. more positive charge for single molecule in dielectric and blue indicates a negative difference, up to $\pm 80$ kJ mol$^{-1}$except for Fig 3.2(b) which uses a scale of $\pm 120$ kJ mol$^{-1}$, see Table 3.3 for maximum and minimum values. Arrangements are shown in Fig 3.1.

itself into structures which are close to tetrahedral. The inclusion of two charges LVa and LVb is an uncommon choice but allows for an explicit dipole.

#### 3.2.1.1 The 7-site charge model

The positions of the charges were decided first varying the oxygen-LVa and oxygen-LVb separation $d_V$ and then varying the oxygen-LT separation $d_{LT}$. It was assumed that $d_T$ should be on the order of the separation from the oxygen to the extra charges in the SAPT-5s model which we shall call $d_0 = 0.322214$ Bohr and for this reason $d_T$ is given as a multiple of $d_0$. Positioning the extra sites closer to the oxygen resulted in very large charges for the oxygen site and the extra sites, though this can perhaps be justified provided the sum of all extra charges and the oxygen charge is close to oxygen's charge in a 3-site model. For the seven site model the charges are given for variations of $d_{LV}$ keeping $d_T$ at $d_0$ in Table 3.4. Table 3.5 shows charges for variations of $d_T$ keeping $d_V = 0.2$ Bohr. In an effort to make some compromise between the quality of the fit and the size of the charges involved, the decisions made were

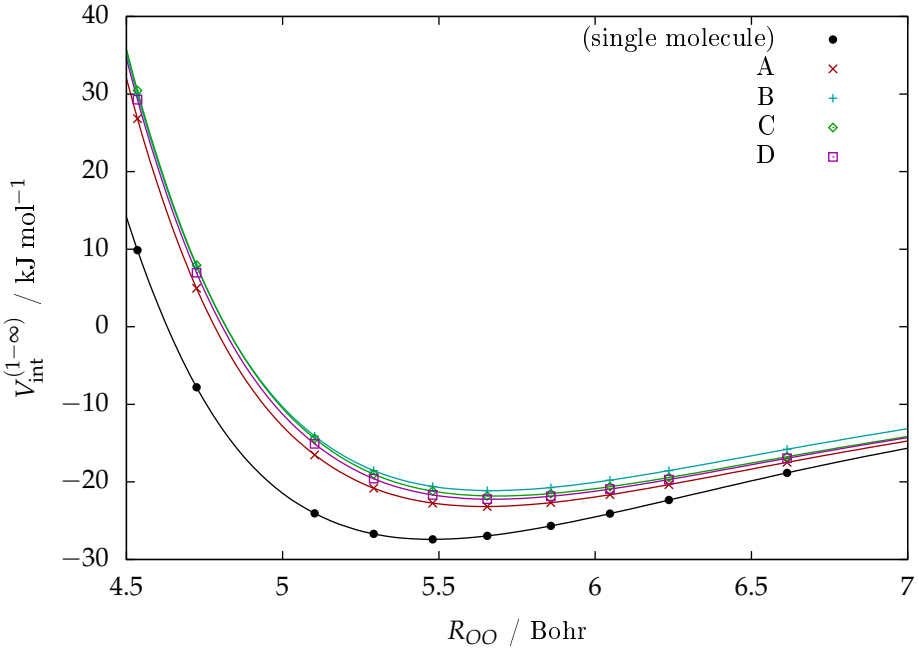



Figure 3.3: Profile plot of $E_{\text{int}}^{(1-\infty)}$ along the global minimum for either a single molecule in dielectric or the central molecule in one of these clusters in dielectric.

$d_V = 0.2$ Bohr and $d_T = d_0 = 0.322214$ Bohr.

| $d_V$ (Bohr) | 0.1 | 0.2 | 0.3 | 0.5 | 0.7 |
|---|---|---|---|---|---|
| $q_O$ | 13.84250 | 7.04763 | 5.80155 | 5.10059 | 4.54034 |
| $q_H$ | 0.41274 | 0.41211 | 0.41129 | 0.41094 | 0.41704 |
| $q_{LT}$ | −2.79804 | −2.80515 | −2.81133 | −2.78128 | −2.59295 |
| $q_{LVa}$ | 3.25267 | 2.76865 | 2.10087 | 1.36545 | 0.95336 |
| $q_{LVb}$ | −12.33923 | −5.03020 | −3.10236 | −1.72537 | −1.14190 |
| Goodness of Fit (kJ mol$^{-1}$) | 0.3569 | 0.3662 | 0.3837 | 0.4539 | 0.5810 |

Table 3.4: Charges (in a.u.) produced using Mulfit to match the ISA multipoles for the charge set-up shown in Fig 3.5, where $d_T$ has been kept at $d_0$.

| $d_T/d_0$ | 0.6 | 0.8 | 1.0 | 1.2 | 1.4 |
|---|---|---|---|---|---|
| $q_O$ | 17.98343 | 10.39964 | 7.04763 | 5.16403 | 4.10320 |
| $q_H$ | 0.42216 | 0.40688 | 0.41211 | 0.41729 | 0.41704 |
| $q_{LT}$ | −8.34391 | −4.53246 | −2.80515 | −1.88151 | −1.33609 |
| $q_{LVa}$ | 5.51138 | 3.81978 | 2.76865 | 2.15133 | 1.68224 |
| $q_{LVb}$ | −7.60934 | −5.96826 | −5.03020 | −4.38692 | −3.95758 |
| Goodness of Fit (kJ mol$^{-1}$) | 0.3438 | 0.3540 | 0.3662 | 0.3800 | 0.3943 |

Table 3.5: Charges (in a.u.) produced using Mulfit for the charge set-up shown in Fig 3.5, where $d_V$ has been kept at 0.2 Bohr.

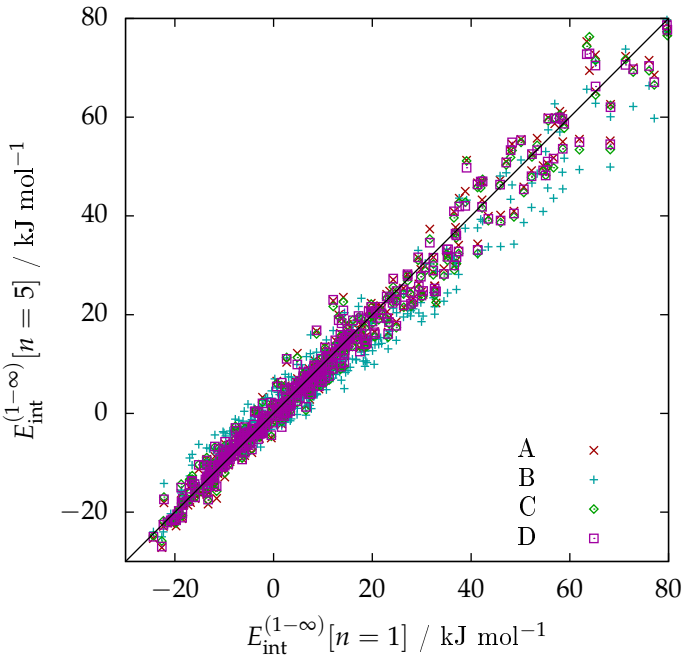



Figure 3.4: Plot of $E_{\text{int}}^{(1-\infty)}$ using charge models fitted to the arrangement shown in Fig 3.1, against $E_{\text{int}}^{(1-\infty)}$ using a multipole model describing a single water molecule in a dielectric.

#### 3.2.1.2 5-site tetrahedral charge model

Given that seven is an unusually large number of sites for a water molecule, a five-site charge model omitting the "LVa" and "LVb" charges was made. Now confident that small changes in $d_{LT}$ would not strongly affect the goodness of the fit, fewer options were tried and $d_T = d_0$ was regarded as an appropriate choice. See Table 3.6.

| $d_{LT}/d_0$ | 0.25 | 1.00 | 4.00 |
|---|---|---|---|
| $q_O$ | −8.93990 | −3.40282 | −0.53854 |
| $q_H$ | 0.54636 | 0.57758 | 0.30705 |
| $q_{LT}$ | 3.92359 | 1.12372 | −0.03778 |
| Goodness of Fit (kJ mol$^{-1}$) | 1.2788 | 1.4611 | 2.2638 |

Table 3.6: Charges (in a.u.) produced using Mulfit for a charge set-up using sites O, H,and LT as shown in Fig 3.5.

#### 3.2.1.3 SAPT-5s site model

A third charge set-up was tried using the site locations of the SAPT-5s model [14]. In this model the extra charges are placed out of the plane, but "below" the oxygen similar to TIP4P models. The SAPT-5s charge locations were also used to produce a charge model with Mulfit. The benefits of doing this rather than lifting the charge model entirely were that it could be fitted to match any charge model made using the DF-ISA method. In particular, it is possible to directly compare ISA multipole models made to describe water in a dielectric without polarisability (see Sec 3.1.4), then use these to give charge-only models and combine them with

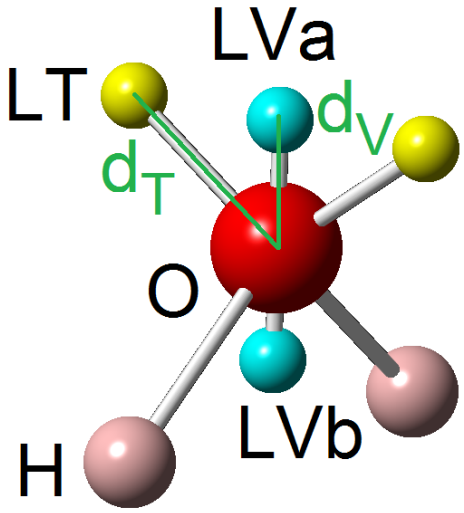



Figure 3.5: Charge arrangement for the seven-site water model. Parameters $d_T$ and $d_V$ were varied, and the charges of all sites were then adjusted using Mulfit. The oxygen and hydrogen sites are positioned according to the vibrationally averaged structure as in the previous section. Angle LT-O-LT is the tetrahedral angle given by $\cos^{-1}(-\frac{1}{3}) \approx 109.47°$.

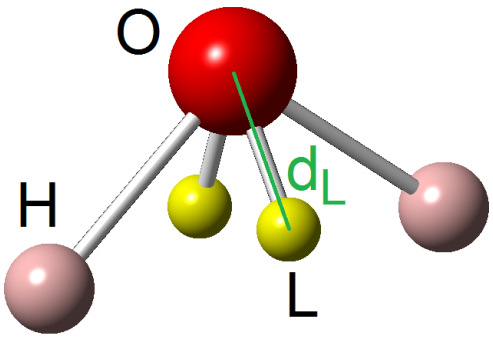



Figure 3.6: Charge arrangement for the SAPT-5s water model water model[14]. Parameters are given in Table 3.7.

this potential. Charge values for both the SAPT-5s model and our model using the SAPT-5s charge locations are given in Table 3.7.

| | SAPT-5s | Mulfit |
|---|---|---|
| $q_O$ | 0.2585046 | 0.11923 |
| $q_H$ | 0.5640504 | 0.52245 |
| $q_{LT}$ | $-0.6933027$ | $-0.58207$ |
| $d_L$ | 1.063 Å | - |
| Angle L-O-L | 58.15 ° | - |

Table 3.7: Parameters for the SAPT-5s charge model[14] and our own model using the same locations. Charge values in a.u.

#### 3.2.1.4 Comparison of charge models

The three charge models mentioned above were compared to the ISA multipole model they are intended to imitate. This can be done by comparing the multipole moments, as well as the total electrostatic energy. The magnitude of the multipole moments for each rank are given in Table 3.9, electrostatic energy is shown is scatter plots in Fig 3.8 with rmse in regions categorised according to SAPT(DFT) total interaction energy given in Table 3.10. All charge models are also compared to the ISA multipole model in Figure 3.7 and Table 3.8 using an isodensity surface. All results indicate that the 7-site model is superior.

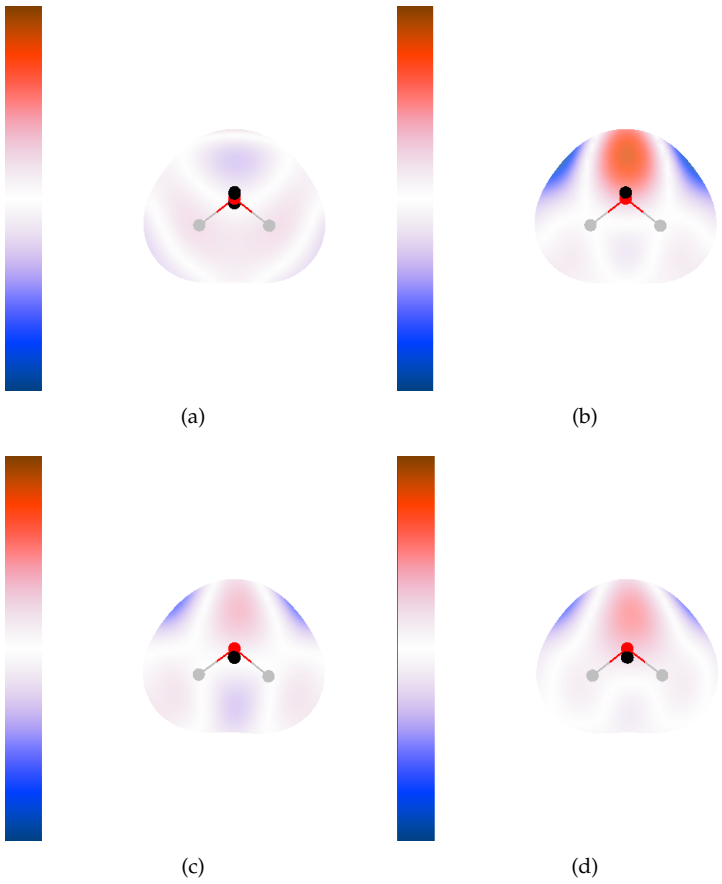



Figure 3.7: Differences in electrostatic energy between the rank 4 multipole model and charge-only models interacting with a unit charge at a surface density of $10^{-2}$ a.u., shown for 3.7(a) the 7-site model, 3.7(b) the five site tetrahedral model, 3.7(c) the SAPT-5s charge model, 3.7(d) charges fitted to the same locations as the SAPT-5s model. Maximum errors are given in Table 3.8. The range is $\pm 50$ kJ mol$^{-1}$, red indicates more positive charge for the charge-only model.

| Charge Model | 7-site | 5-site tetrahedral | SAPT-5s charges | SAPT-5s sites |
|---|---|---|---|---|
| Max. dif. | 12.421 | 40.625 | 18.572 | 24.285 |
| Min. dif. | $-15.800$ | $-47.237$ | $-29.895$ | $-30.104$ |
| rmse | 7.021 | 18.047 | 10.970 | 11.010 |

Table 3.8: Differences in kJ mol$^{-1}$ between the rank 4 multipole model for water and charge-only models at a surface density of $10^{-2}$ a.u. The surface is shown in Fig 3.7. Positive difference indicates more positive charge for the charge-only model.

| Multipole Moment Rank | ISA | 7 sites | 5 site tetrahedral | SAPT-5s sites |
|---|---|---|---|---|
| 1 | 0.741 | 0.740 | 0.745 | 0.750 |
| 2 | 2.173 | 2.096 | 1.894 | 2.117 |
| 3 | 4.694 | 3.996 | 5.943 | 5.770 |
| 4 | 8.949 | 7.118 | 9.964 | 9.751 |

Table 3.9: Magnitude of total multipole moments for the water molecule in a.u. for each rank for each charge model used.

| Charge setup | 7 sites | 5 site tetrahedral | SAPT-5s sites |
|---|---|---|---|
| $w > 1.4$ | 0.607 | 1.662 | 0.838 |
| $> 1, < 1.4$ | 0.402 | 1.698 | 0.679 |
| $> 0.4, < 1$ | 0.579 | 2.043 | 0.879 |
| $> 0.1 < 0.4$ | 1.223 | 4.075 | 2.301 |
| $> 0.001 < 0.1$ | 2.445 | 6.748 | 3.381 |

Table 3.10: rmse in kJ mol$^{-1}$ for different charge models compared with the ISA multipole model. $w$ as in Table 2.6.

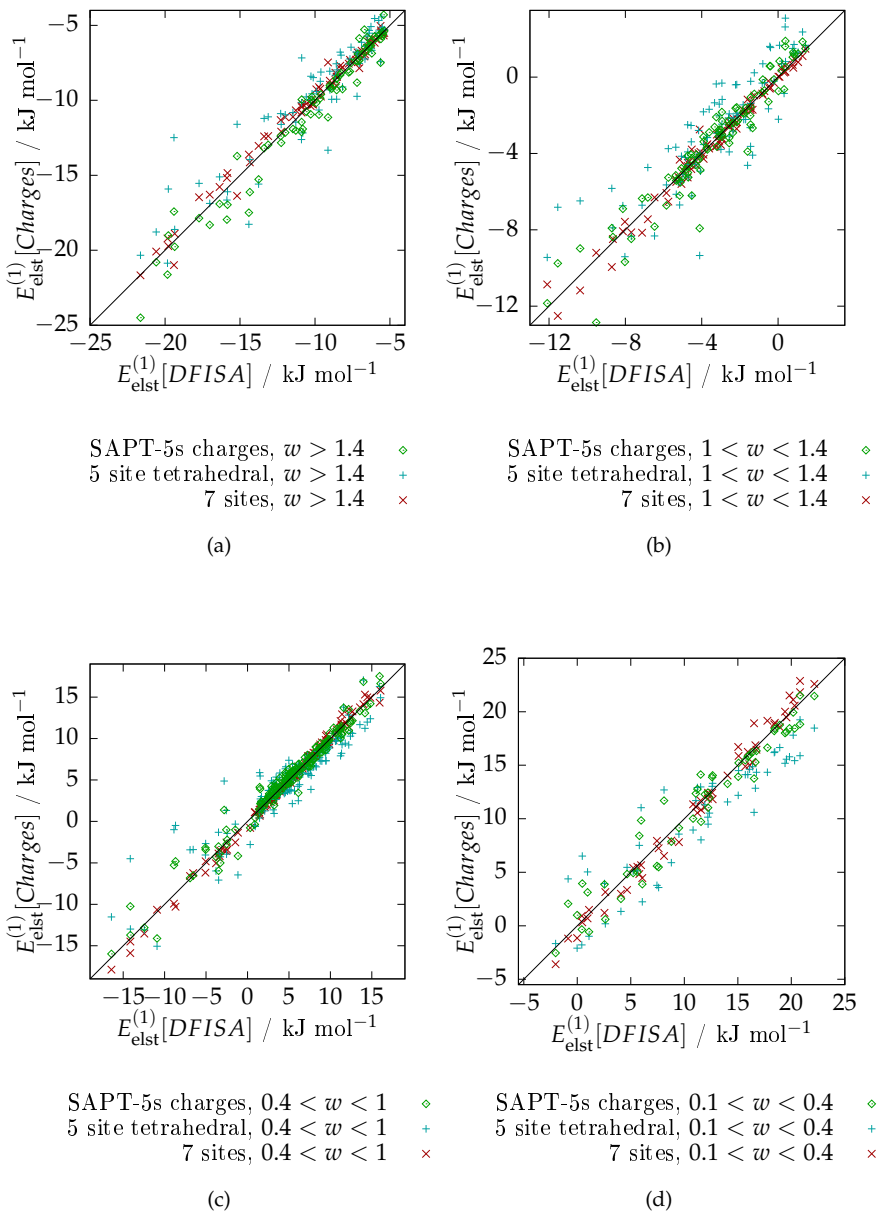



Figure 3.8: Scatter plots of $E^{(1)}_{\text{elst}}[DFISA]$ against $E^{(1)}_{\text{elst}}[Charges]$ using the seven-site, five-site tetrahedral and SAPT-5s site locations. Plots are for different values of $w$, i.e. different energy ranges, see Table 2.6.

### 3.2.2 Polarisation Model

At first it was assumed that the polarisability should be scaled to match the full anisotropic rank 3 model. For this purpose, each dimer configuration from the set of 2048 was given the same statistical weight as in the final stage of the anisotropic fit, and a single prefactor $s$ was chosen accordingly, i.e. the polarisability was chosen as $\alpha^{=}\alpha_0^a$ where $\alpha_0^a = \frac{1}{3}(\alpha^a_{10,10} + \alpha^a_{11c,11c} + \alpha^a_{11s,11s})$ for each atom $a$. Using this method, $s = 1.208$, suggesting at second order approximately five sixths of the polarisation energy is accounted for.

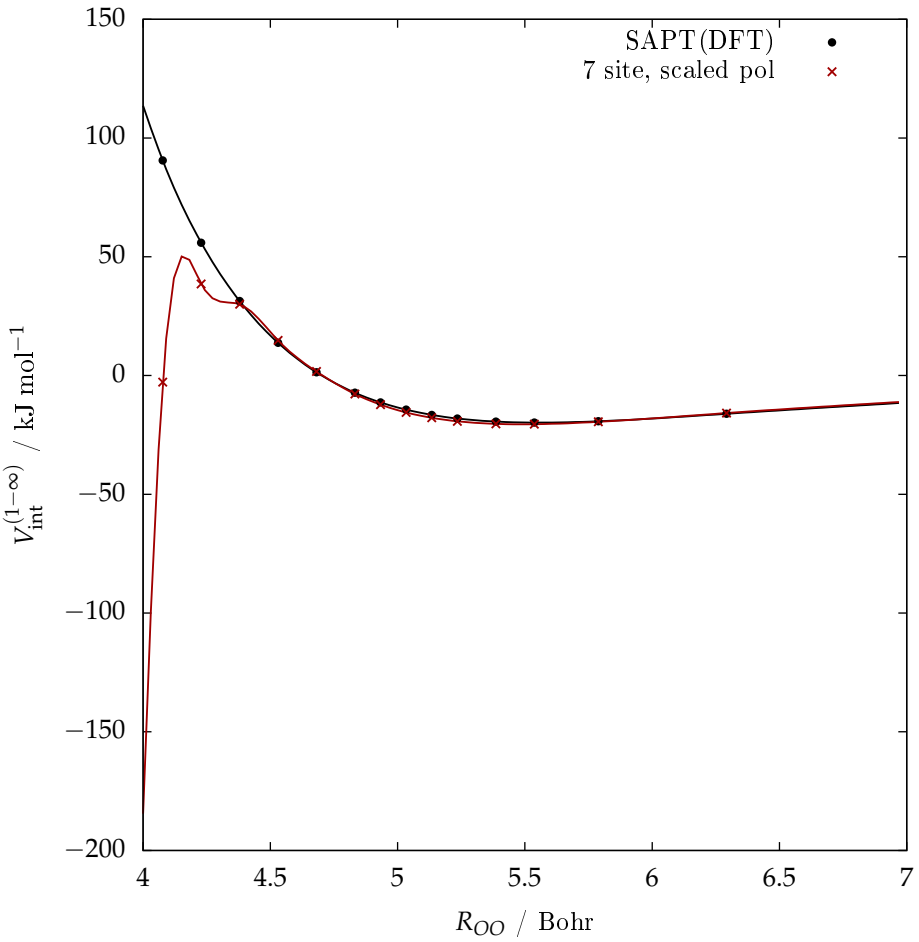
Regrettably, when this polarisability was used the potential fell victim to the "polarisation catastrophe", i.e. the short-range potential was unable to overcome the undamped polarisation energy and the total energy became extremely negative below a given separation, in this case around 4 Bohr. A demonstration of this is given in Fig 3.9(a), which plots the points for one of the profiles described in Sec 3.2.3. There is nothing instantly notable about the points for which the polarisation fails; the most severe example is the one plotted and is shown in Fig 3.9(b).

To resolve this it was necessary to reduce the polarisability. There is no clear-cut method of determining at what stage it is safe to say that the polarisation energy will always be sufficiently small. Other water models have used reduced polarisabilties for this purpose, with SWM4-DP [123] and SWM4-NDP [124] both having polarisability densities of around 1 Å$^3$ the SWM6 [125] six site model had a polarisability density of only 0.88 Å$^3$ (compared with the accepted value of 1.44 Å$^3$), but it is obviously desirable to have as little reduction as possible. Even in cases where no fit point indicated failure of the interaction energy to remain positive at very short separations, minimisation of hexamer structures sometimes still failed. Eventually it was found that, using either 5-site charge arrangement, the polarisability didn't cause any problems if it was scaled by $s = 0.9$. Comparing this to the previously mentioned value of $s = 1.208$ suggests that the polarisation energy has been reduced to around three quarters the value it had when using the rank 3 anisotropic damped model. The resulting polarisability is shown in Table 3.11. The sum of polarisability densities from this model is 1.28 Å$^3$.

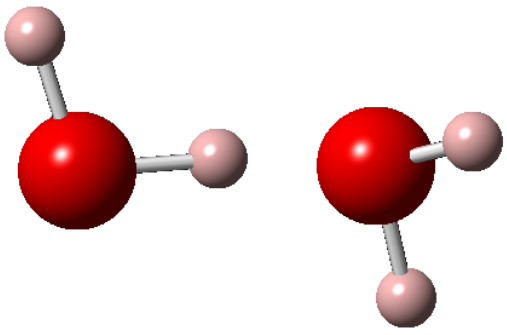
| Atom | Polarisability | Average Polarisability of L3 model |
|---|---|---|
| O | 5.40801522 | 6.0089058 |
| H | 1.60741755 | 1.7860195 |

Table 3.11: Distributed polarisabilities in atomic units for the model chosen, and compared with the isotropic part of the damped L3 anistropic model.

In our own case, the seven site model also failed for much smaller polarisabilities than the five site models. For this reason the seven site model, which gave by far the best electrostatic energy, was not used.




(a)



(b)

Figure 3.9: Profile plot demonstrating the problem of undamped polarisability with the orientation of the molecules as shown.

### 3.2.3 Short-range Energy

The short-range energy was initially fitted to the same set of 2048 points as had been used in fitting the anisotropic potential, however the limitations of an isotropic potential made it difficult to find a reasonable fit to the global minimum profile using those points only. As an alternative, a set of dimers which appeared in the five lowest-lying hexamers was used. This set consisted of 15 points taken for each of the pairs between which lines are drawn in Fig 3.10, i.e. those for which hydrogen bonds may occur. The intent was that such a set of configurations should lend itself to accurate hexamer energies and structures, but also to the dimer minimum since many of these pairs resemble the minimum energy dimer; ring structures especially tend to consist of chains of dimers very close to the minimum configuration.

It could be argued that this set favours certain structures undesirably because, for instance, there are eight pairs taken from the cage structure but only six pairs taken from the ring structure. The obvious solution would be to include all pairs, i.e. 15 for each structure, but this would introduce orientations at separations for which molecules barely interact at all. So, we concede to stick with the pairs indicated, which would be those which appear in the hexamers within an arbitrary cutoff distance of about 3 Å.

Excepting cases which failed due to polarisation problems as described in Sec 3.2.2, short-range fits were carried out using both five site multipole models. Both models reproduced the global minimum and other points in the fit similarly well, however when use to find clusters either through basin hopping searches or reminimisation differences between the models surfaced: The model using SAPT-5s charges reproduced the geometric parameters but had poor energetic ordering; the model using a tetrahedral charge arrangement had better ordering but gave a "flat" pentamer and tetramer ring.

#### 3.2.3.1 Fitting the short-range energy using a tetrahedral charge arrangement

It was not easy to select which anchor strength and weighting parameter led to the best potential. Starting with the weighting parameter $E_0 = 10$ kJ mol$^{-1}$ as for the anisotropic potential, anchors strengths were varied giving rmse as shown in 3.14. It is worth noting that these rmse use the new set of geometries, for which a much larger fraction of the points used fall into low-energy categories. For this reason we see that weaker anchors give higher rmse at higher energies. Because of this, the strong anchors ($A = 0.1$) were preferred; these gave slightly higher rmse for the lowest energy categories (0.818 vs 0.779 kJ mol$^{-1}$ and 1.782 vs. 1.686 kJ mol$^{-1}$) and lower rmse at higher, but still negative, energies (2.780 vs 2.913 kJ mol$^{-1}$). Changing the weighting parameter $E_0$ made little difference on the rmse in low energy categories; in the lowest category (total energies below 6.727 kJ mol$^{-1}$) rmse varied from 0.817 kJ mol$^{-1}$ at $E_0 = 5$ kJ mol$^{-1}$to 0.847 kJ mol$^{-1}$ at $E_0 = 25$ kJ mol$^{-1}$, see Table 3.13.

Because the rmse did not give easy choices; and because there was concern using an isotropic model would limit the ability to reproduce accurate clusters, the decision on $A$ and $E_0$ was not made until the dimer and hexamer structures had been looked at. The dimer structures were found using a basin-hopping search, which was ran at 500K and allowed up to 100 steps (although only one minimum was found). The hexamers were found by reminimising results of a previous basin-hopping search carried out for the anisotropic potential.

Dimer parameters (as shown in Fig 3.11) are given for $E_0 = 5$ and $E_0 = 15$ kJ mol$^{-1}$, and

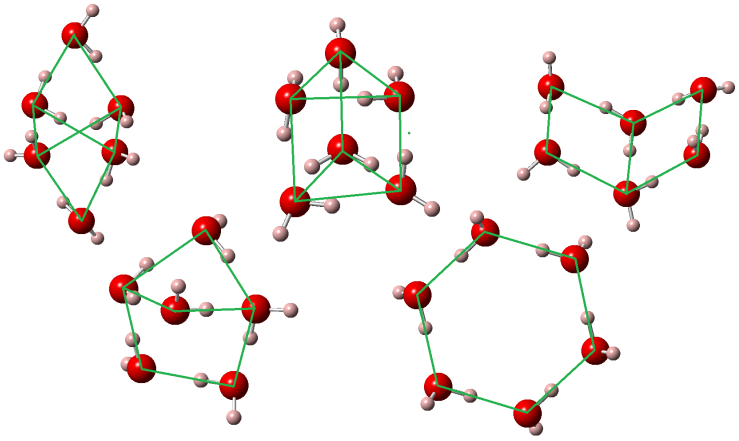


Figure 3.10: Pairs from hexamers used in fitting the short-range energy, indicated by green lines. These were taken from the anisotropic model.

for $A = 0.1$ and $A = 0.001$ in Table 3.12. The separation $R_{OO}$ varied by only 0.02 Å across all fits. This is very small, for example it is less than the difference between separations using SAPT and CCSD(T) methods. The small angle, $\Theta$, should be around 5° but varies from -2.20° to 1.88°. In this case the fits made using $A = 0.1$, for which $\Theta = 1.86°$ and 1.88° for $E_0 = 15$ and $E_0 = 5$ kJ mol$^{-1}$respectively, appear to be the better choice. These gave the large angle $\Phi = 119.81°$ and 119.59° when we would like a value of around 124°. Using $A = 0.001$ and $E_0 = 15$kJ mol$^{-1}$gives $\Phi = 128.51°$, which is only very slightly closer to the CCSD(T) value of 124.92°. Of the two models using $A = 0.1$, the minimum depth is closer by around 0.3 kJ mol$^{-1}$ for $E_0 = 5$ kJ mol$^{-1}$, although both are within 1 kJ mol$^{-1}$ of the reference energy and the well-depth according to SAPT and CCSD(T) calculations.

| | $R_{OO}$ Å$^{-1}$ | $\Theta$ (°) | $\Phi$ (°) | $E_{\text{int}}^{(1-\infty)}$ / kJ mol$^{-1}$ |
|---|---|---|---|---|
| $E_0 = 15$ kJ mol$^{-1}$,$A = 0.1$ | 2.901 | 1.86 | 119.81 | −20.089315 |
| $E_0 = 15$ kJ mol$^{-1}$,$A = 0.001$ | 2.895 | 0.33 | 128.51 | −20.246428 |
| $E_0 = 5$ kJ mol$^{-1}$,$A = 0.1$ | 2.901 | 1.88 | 119.59 | −20.409838 |
| $E_0 = 5$ kJ mol$^{-1}$,$A = 0.001$ | 2.881 | -2.2 | 155.05 | −20.731687 |
| Anisotropic model | 2.917 | 5.26 | 126.33 | −21.013 |
| CC-Pol[a] | 2.911 | 6.48 | 122.66 | −22.179 |
| SAPT[b] | 2.953 | 4.6 | 124 | −20.334 |
| CCSD(T)[c] | 2.909 | 4.47 | 124.92 | −21.087 |

Table 3.12: Parameters of the global minimum using different values of $E_0$ and $A$ for the tetrahedral water model, as shown in Fig 3.11. *a* From [130]. *b* From [14]. *c* Geometry optimised in TZ2P(f,d)+dif basis set, energy from MP2 CBS extrapolation. From [131].

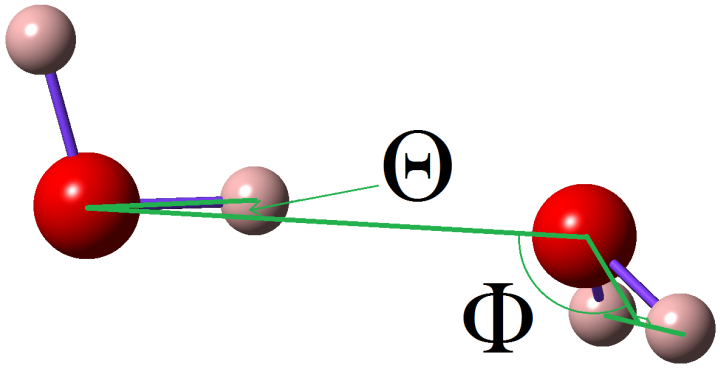



Figure 3.11: The minimum-energy configuration water dimer. The O-O separation and measurements for the angles indicated for different fit of the tetrahedral water model are given in Table 3.12

.

#### 3.2.3.2 Evaluation of fit using a tetrahedral charge arrangement

For both fits with $A = 0.1$, the energetic ordering of hexamers given in Table 3.15 is as follows: Cage, prism, bag, book, cup, ring (the hexamers are shown in Fig 4.8). The accepted order is prism, cage, book, bag, ring, cup. The prism and cage are within 1 kJ mol$^{-1}$of each other and it is common for water models to order them the wrong way round, more notable fault is the placing of the bag before the book and the cup before the ring. Given that the prism structures are similar for the isotropic and anisotropic models, comparing the relative energies suggests that the ring structure is similar but that the book, bag and cup structures are not. In fact the book structure in the isotropic model is much "flatter"; the O-O-O angles which link the "spine" of the book to its corners are 165.0° and 146.2° compared with 114.5° and 112.1°.

The hexamer energies using the isotropic fits also span a smaller range than the CCSD(T) calculations. For example the ring hexamer has an energy around 5.56 kJ mol$^{-1}$higher than the prism using the $E_0 = 5$kJ mol$^{-1}$, $A = 0.1$ fit, whereas for CCSD(T) optimisations this difference can be as high as 7.66 kJ mol$^{-1}$.

The model fails to reproduce the ring structures for the pentamer and tetramer. These structures are expected to form a ring with molecules in a chiral arrangement, but instead the rings formed are "flat", i.e. all atoms are coplanar, see Fig 3.13. This failure occurs regardless of fitting parameters and may be due to the charge model: Water rings are composed of a chain of hydrogen bonds, where the positive oxygen site of one molecule is attracted to a hydrogen atom in another molecule. For our tetrahedral charge model, it is these two regions in particular where the electrostatic energy is reproduced poorly even compared to using only the ISA charges, see Fig 3.12.

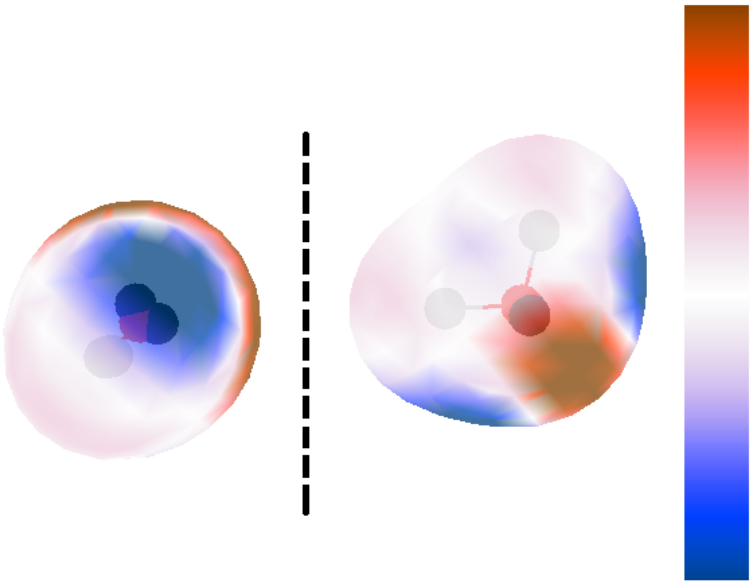
This draws attention to a limitation of using Mulfit to decide charge values: There is no convenient way of biasing the fit to work well in a given area. Consequently we turn to another charge model.

| $E_0$ /kJ mol$^{-1}$ | 5 | 10 | 15 | 25 |
|---|---|---|---|---|
| $w > 1.4$ | 0.817 | 0.818 | 0.834 | 0.847 |
| $> 1, < 1.4$ | 1.857 | 1.782 | 1.782 | 1.783 |
| $> 0.4, < 1$ | 3.445 | 2.780 | 2.639 | 2.597 |
| $> 0.1, < 0.4$ | 6.514 | 3.863 | 3.165 | 2.969 |
| $< 0.1$ | 11.093 | 11.439 | 12.959 | 12.916 |

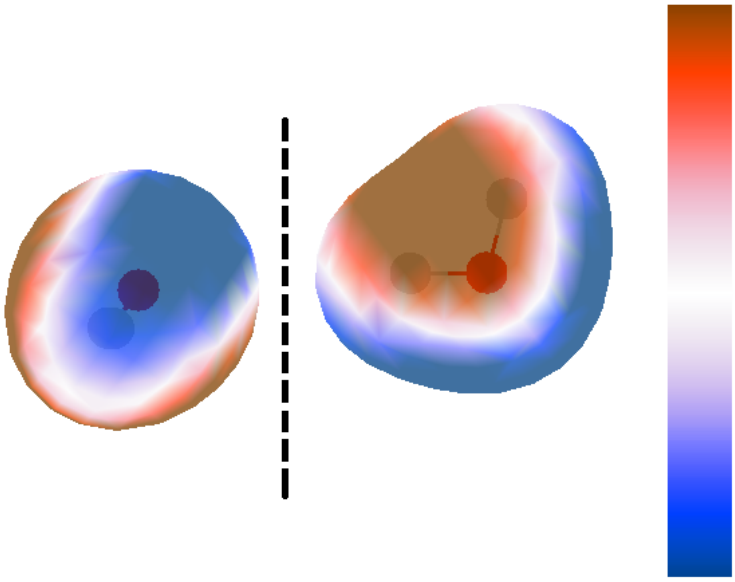
Table 3.13: rmse in kJ mol$^{-1}$ from varying the weighting parameter $E_0$, with $A = 0.1$, for water with a tetrahedral charge arrangement. $w$ as in Table 2.6.

| A | 0.001 | 0.01 | 0.1 |
|---|---|---|---|
| $w > 1.4$ | 0.778 | 0.779 | 0.818 |
| $> 1, < 1.4$ | 1.672 | 1.686 | 1.782 |
| $> 0.4, < 1$ | 2.931 | 2.913 | 2.780 |
| $> 0.1, < 0.4$ | 4.315 | 4.222 | 3.863 |
| $< 0.1$ | 10.450 | 10.789 | 11.439 |

Table 3.14: rmse in kJ mol$^{-1}$ from varying the anchor strength $A$, with $E_0 = 10$ kJ mol$^{-1}$, for water with a tetrahedral charge arrangement. $A = 0.1$ was chosen. $w$ as in Table 2.6.



(a)



(b)

Figure 3.12: The difference in electrostatic energy between the L4 ISA multipole model and a charge model for a unit charge on a surface of electron density of $10^{-2}$ a.u, arranged to indicate the hydrogen bond present in the minimum energy dimer or in small rings. Shown for 3.12(a) the 5-site tetrahedral charge model, and 3.12(b) the charges from the rank-3 ISA model. The range is $\pm 30$ kJ mol$^{-1}$, red indicates more positive charge for the charge-only model. The tetrahedral model performs particularly poorly along the line of the global minimum.

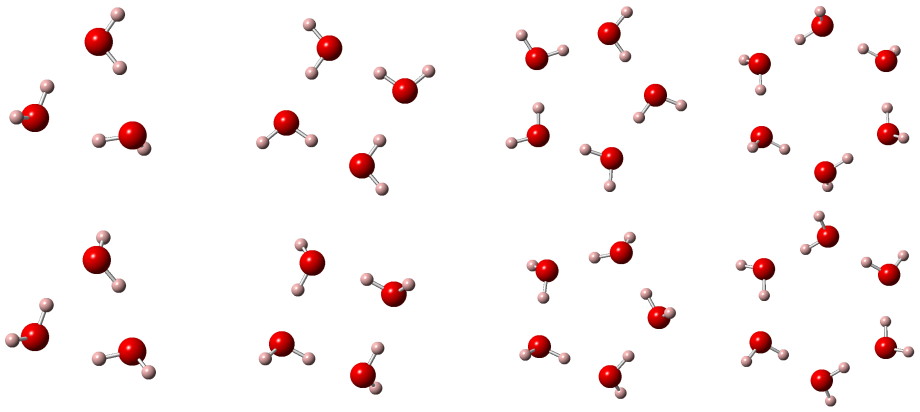


Figure 3.13: Comparison of rings structures of up to 6 molecules using the tetrahedral charge arrangements (top) and anisotropic potential (bottom). For tetramer and pentamer rings the tetrahedral model incorrectly gives a flat ring.

| Structure | Anisotropic | $E_0 = 5$ kJ mol$^{-1}$ | $E_0 = 15$ kJ mol$^{-1}$ |
|---|---|---|---|
| Prism | −195.569 | −188.965 | −186.545 |
| Cage | −195.726 | −187.910 | −185.732 |
| Book | −193.956 | −186.002 | −184.760 |
| Bag | −191.604 | −186.732 | −185.318 |
| Ring | −189.939 | −183.401 | −182.9884 |
| Cup | −187.211 | −183.936 | −183.006 |

Table 3.15: Energies of low-lying hexamers for different fits of the tetrahedral isotropic water potential, and the anisotropic potential it is intended to match.

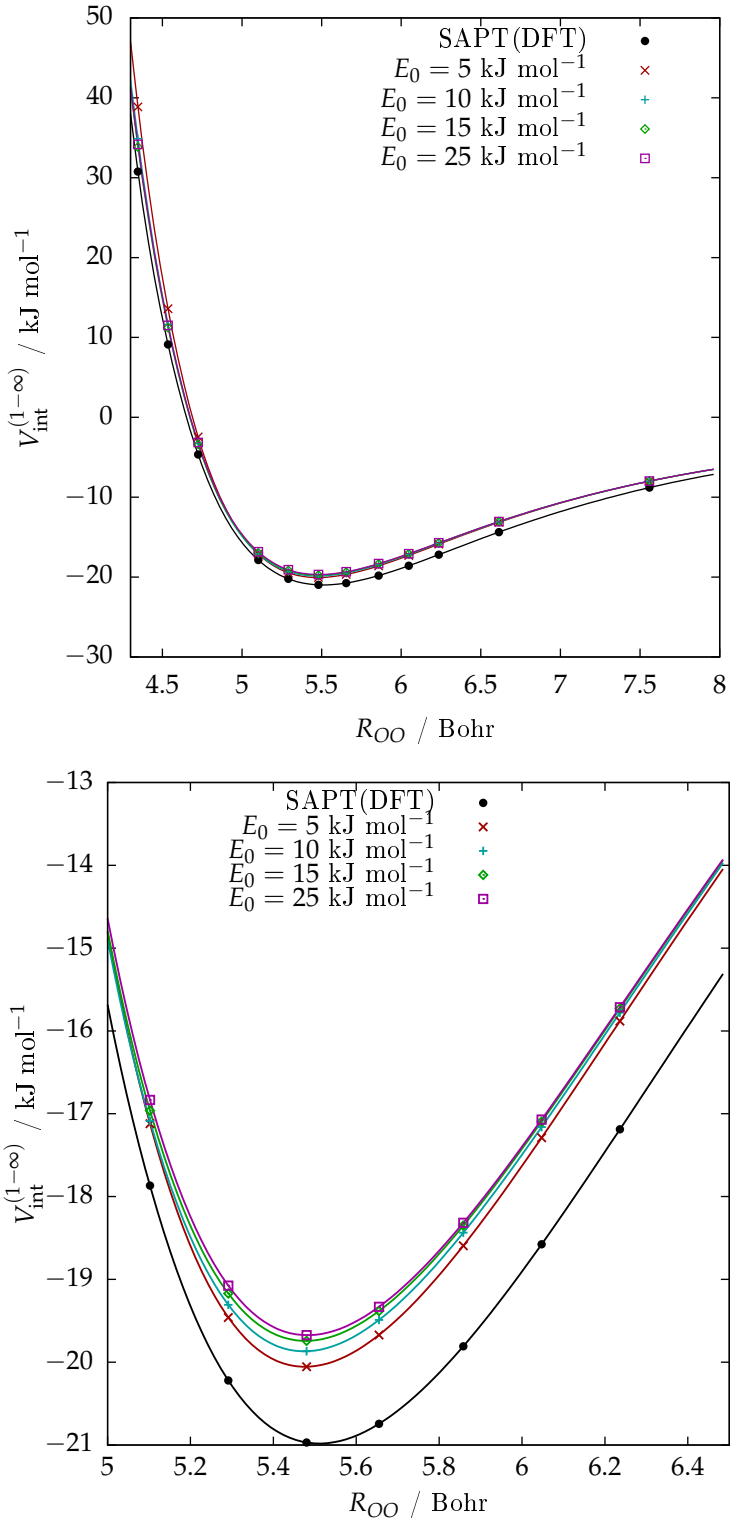



Figure 3.14: Profile plot of $V_{\text{int}}^{(1-\infty)}$ against oxygen-oxygen separation $R_{OO}$, varying weighting parameter $E_0$ with $A = 0.1$, for water with a tetrahedral charge arrangement. $E_0 = 10$ kJ mol$^{-1}$ was chosen.

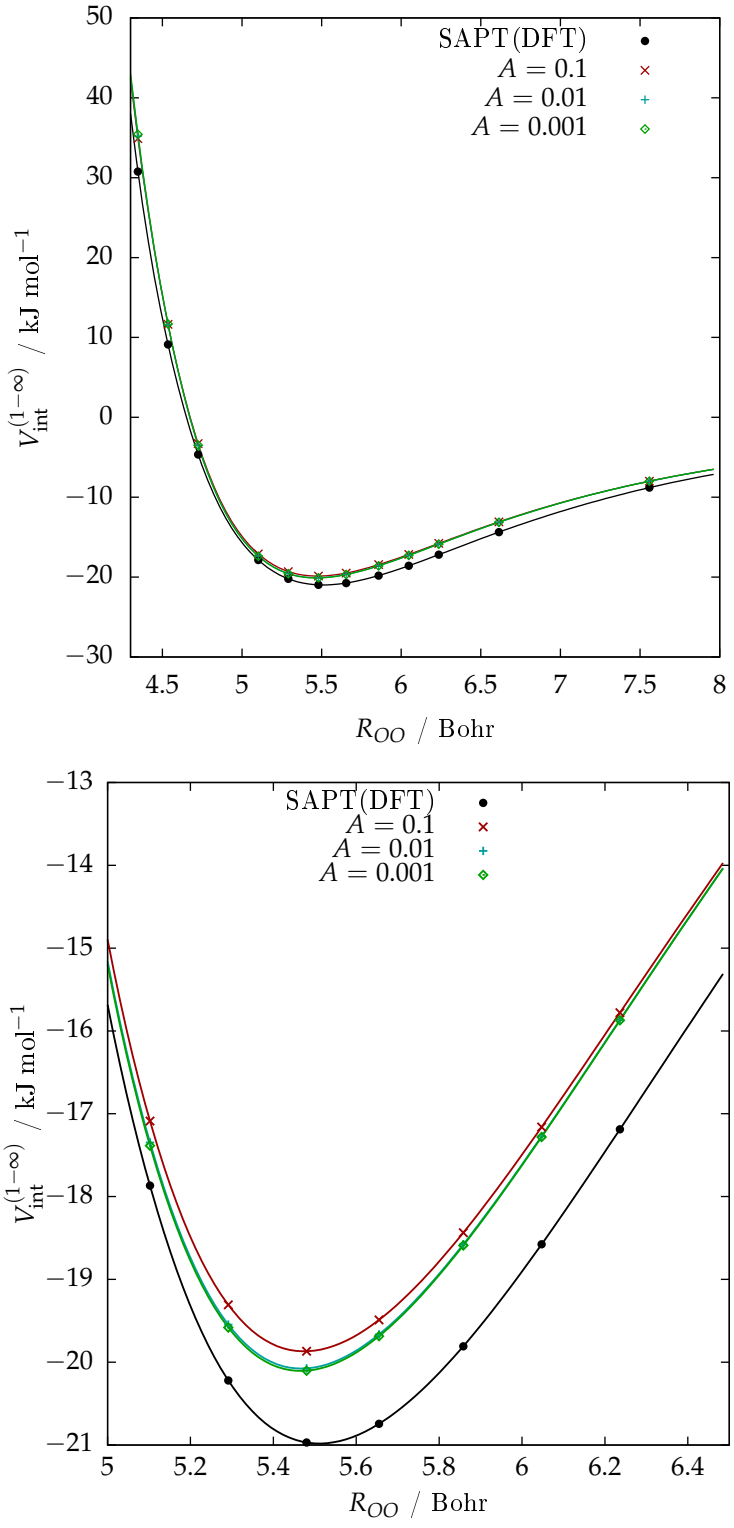



Figure 3.15: Profile plot of $V_{\text{int}}^{(1-\infty)}$ against oxygen-oxygen separation $R_{OO}$, varying weighting parameter $A = 0.1$ with $E_0 = 10$ kJ mol$^{-1}$, for water with a tetrahedral charge arrangement. $A = 0.1$ was chosen.

### 3.2.4 Fitting the short-range energy using the SAPT-5s site locations

The charge model using the SAPT-5s site locations, as described in Sec. 3.2.1, does not suffer from a poor fit in the regions typically involved in hydrogen bonding. Additionally, the SAPT-5s potential has a similar form to our own potential. These factors contributed to the decision to fit a model using this arrangement.

Neither the rmse for negative-energy categories nor the lowest-energy point varied by more than 0.1 kJ mol$^{-1}$ when choosing the weighting parameter for this model provided $E_0 < 50$ kJ mol$^{-1}$, and $E_0 = 25$ kJ mol$^{-1}$ was chosen only because of these values it gave the closest agreement to SAPT(DFT) reference energies at the minimum (it underbinds by around 0.6 kJ mol$^{-1}$).

Reducing the anchor strength from $A = 0.1$ to $A = 0.01$ gave a reduction in rmse for all energy categories except the lowest one, for which rmse increased but only by 0.023 kJ mol$^{-1}$, and also allowed a very slight improvement of the energy at the global minimum (0.06 kJ mol$^{-1}$). The resulting parameters are given in Table 3.18.

| $E_0$ kJ mol$^{-1}$ | 10 | 15 | 25 | 50 |
|---|---|---|---|---|
| $w > 1.4$ | 0.647 | 0.647 | 0.643 | 0.801 |
| $> 1, < 1.4$ | 1.378 | 1.365 | 1.354 | 1.364 |
| $> 0.4, < 1$ | 2.050 | 1.980 | 2.007 | 1.873 |
| $> 0.1 < 0.4$ | 2.777 | 2.391 | 2.356 | 2.773 |
| $< 0.1$ | 21.364 | 22.307 | 19.999 | 11.737 |

Table 3.16: rmse in kJ mol$^{-1}$ for different values of weighting parameter $E_0$, keeping anchor strength at $A = 0.1$, for water using the SAPT-5s site locations. $w$ as in Table 2.6.

| $A$ | 0.1 | 0.01 | 0.001 |
|---|---|---|---|
| $w > 1.4$ | 0.643 | 0.666 | 0.672 |
| $> 1, < 1.4$ | 1.354 | 1.245 | 1.240 |
| $> 0.4, < 1$ | 2.007 | 1.883 | 1.873 |
| $> 0.1 < 0.4$ | 2.356 | 2.346 | 2.349 |
| $< 0.1$ | 19.999 | 17.400 | 17.064 |

Table 3.17: rmse in kJ mol$^{-1}$ for different values of anchors strength $A$, keeping $E_0 = 25$ kJ mol$^{-1}$, for water using the SAPT-5s site locations. $w$ as in Table 2.6.

| Short-range parameters | | |
|---|---|---|
| Pair | $\alpha_{ab}$ | $\rho_{ab}$ |
| O-O | 1.844988 | 6.012939 |
| O-H | 2.269041 | 4.064269 |
| H-H | 1.987949 | 4.314218 |

Table 3.18: Short-range potential parameters in a.u. for the water model fitted using the SAPT-5s charge locations to replicate the ISA multipole model.

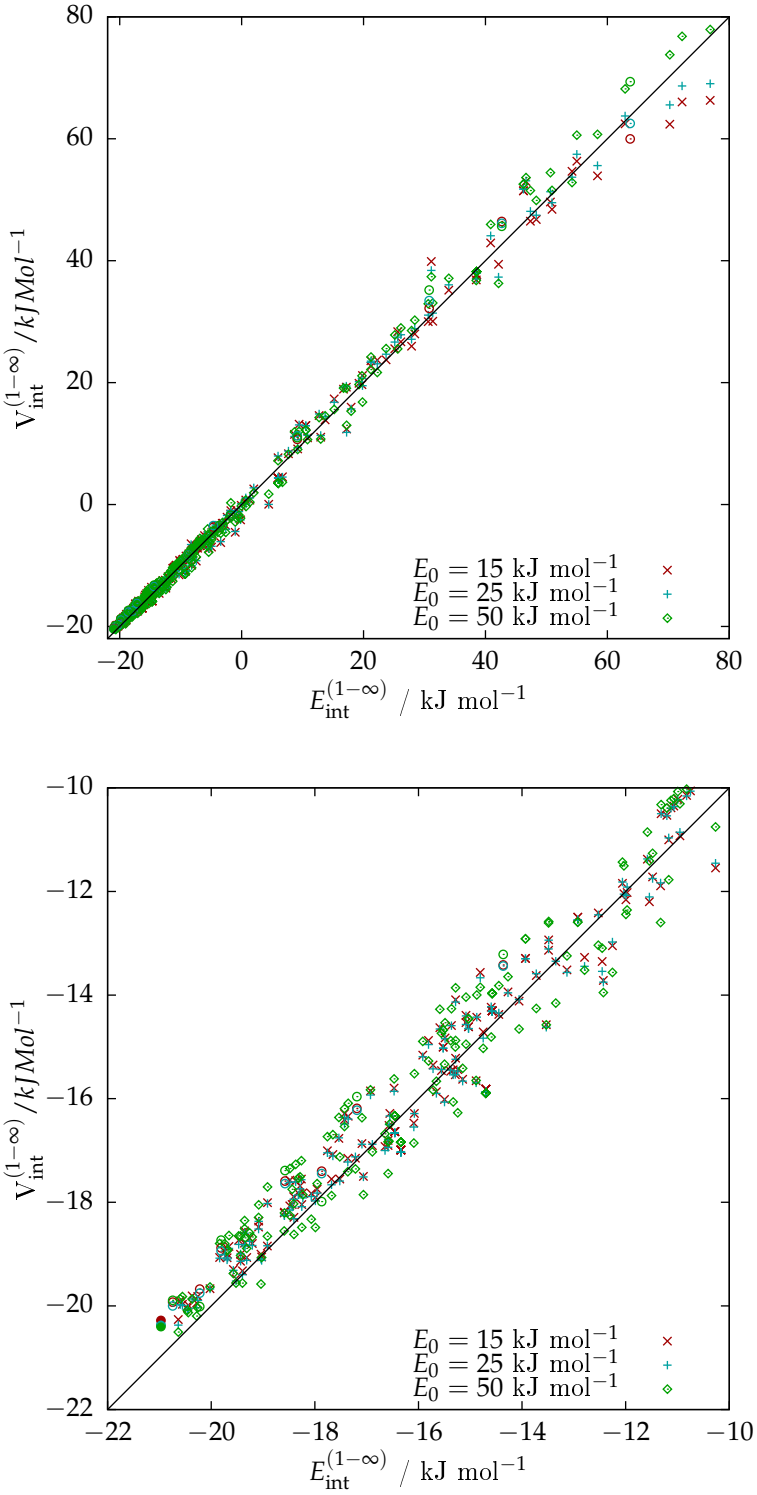



Figure 3.16: Scatter plots of $V_{\mathrm{int}}^{(1-\infty)}$ against $E_{\mathrm{int}}^{(1-\infty)}$ varying weighting parameter $E_0$ when fitting water (using the SAPT-5s site locations) to the infinite-order energy with $A = 0.1$. Plots are over different energy ranges. $E_0 = 25$ kJ mol$^{-1}$ was chosen.

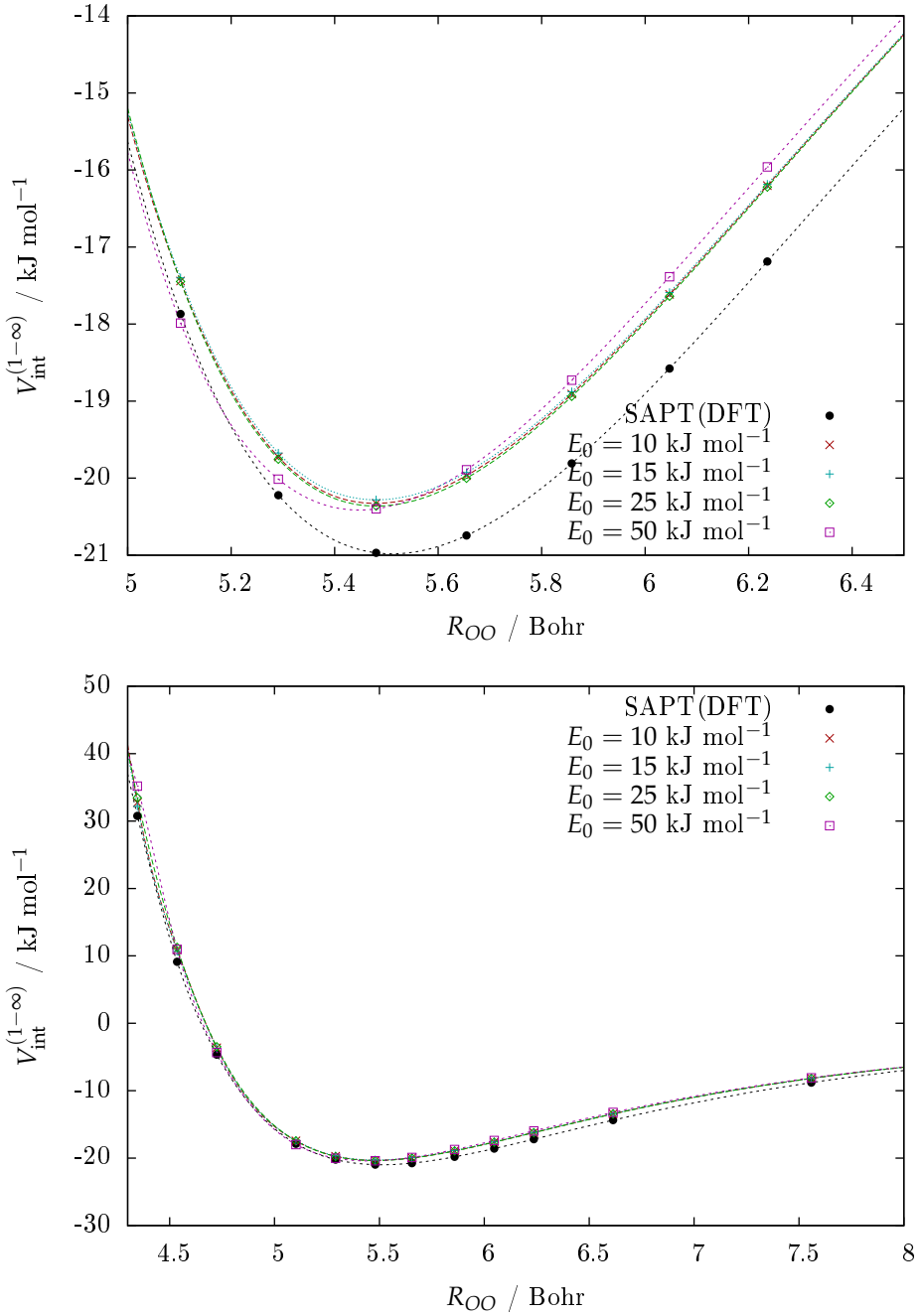



Figure 3.17: Profile plot of $V_{\text{int}}^{(1-\infty)}$ against oxygen-oxygen separation $R_{OO}$, varying weighting parameter $E_0$ when fitting water (using the SAPT-5s site locations) to the infinite-order energy with $A = 0.1$. $E_0 = 25$ kJ mol$^{-1}$ was chosen.

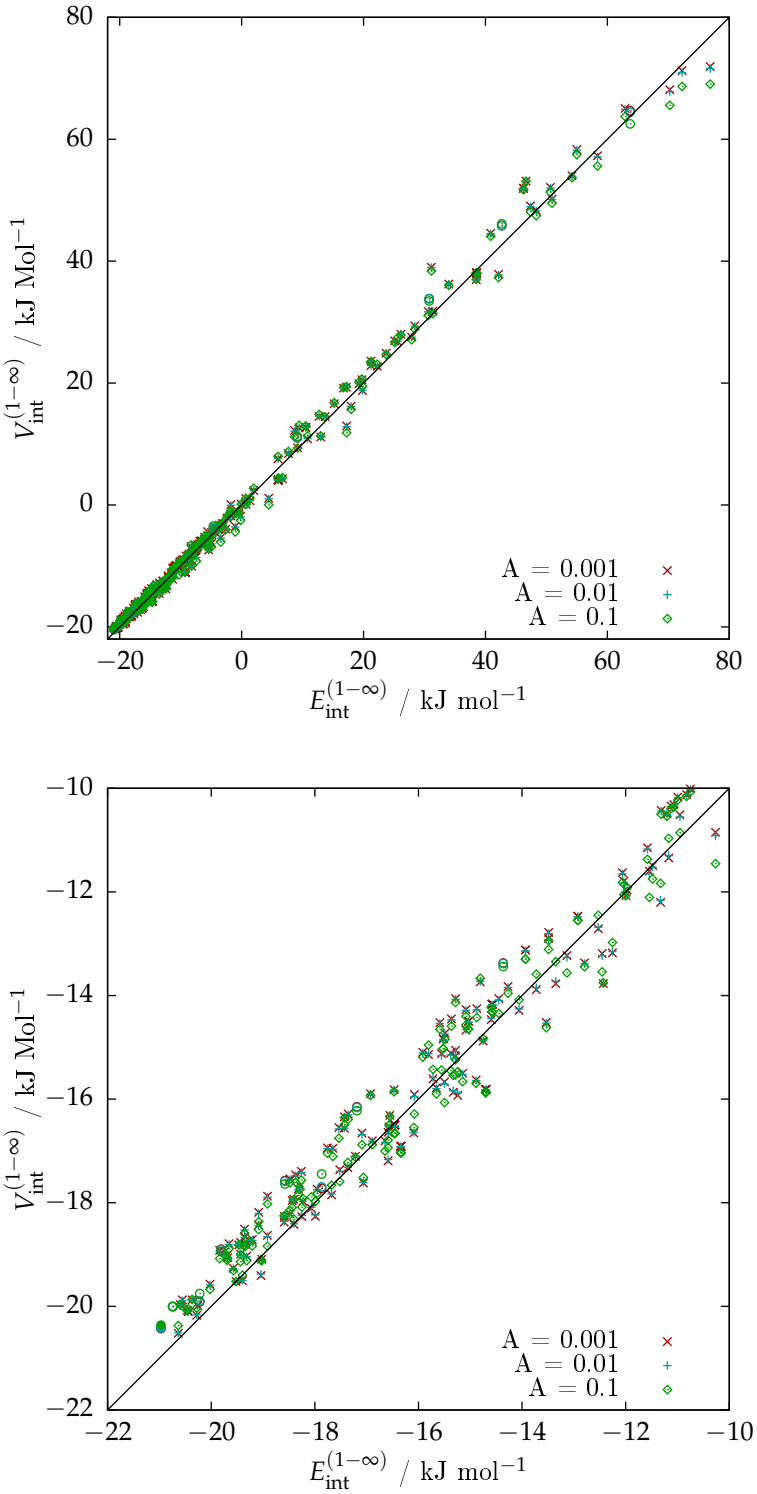



Figure 3.18: Scatter plots of $V_{\text{int}}^{(1-\infty)}$ against $E_{\text{int}}^{(1-\infty)}$ varying anchor strength $A$ when fitting water (using the SAPT-5s site locations) to the infinite-order energy with weighting parameter $E_0 = 25$ kJ mol$^{-1}$. Plots are over different energy ranges. $A = 0.01$ was chosen.

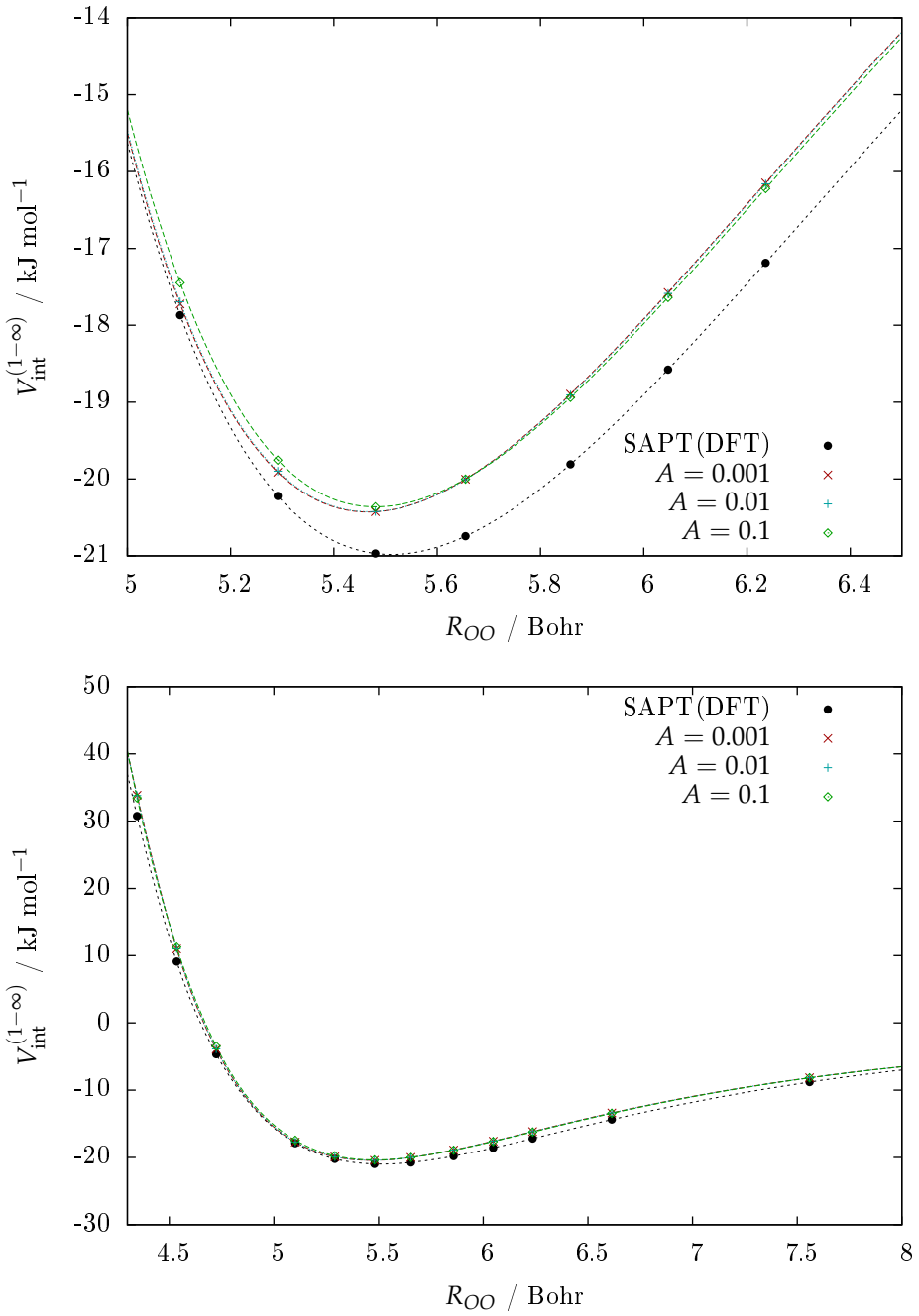



Figure 3.19: Profile plot of $V_{\text{int}}^{(1-\infty)}$ against oxygen-oxygen separation $R_{OO}$, varying anchor strength $A$ when fitting water (using the SAPT-5s site locations) to the infinite-order energy with weighting parameter $E_0 = 25$ kJ mol$^{-1}$. $A = 0.01$ was chosen.

## 3.3 Simplifying the Methane Potential

The steps taken to simplify the methane potential were relatively straight-forward. The charges were taken directly from the ISA multipole model. At first the polarisability was taken directly from the rank 3 anisotropic polarisation model, i.e. $\alpha = \frac{1}{3}(\alpha_{10,10} + \alpha_{11c,11c} + \alpha_{11s,11s})$ was used, but it was found that there was no convincing reason to keep the methane model polarisable at all. As the $C_6$ dispersion model was already fitted to the anisotropic potential directly, there was less motivation to keep the $C_6$ coefficient fixed, and it was also allowed to vary.

### 3.3.1 Charge Model

For methane the ISA charges were assumed to be an acceptable choice. In Fig 3.20 and Table 3.19 the electrostatic potential produced by the charges $E^{(1)}_{\text{elst}}[Q]$ is compared to the multipole model $V^{(1)}_{\text{elst}}$. Fig 3.21 plots $E^{(1)}_{\text{elst}}[Q]$ against $E^{(1)}_{\text{elst}}[DM]$. Clearly there is a tendency for the electrostatic energy using this model to be too small, but as the electrostatic energy is typically small compared with the dispersion and exchange energies this was not too concerning.

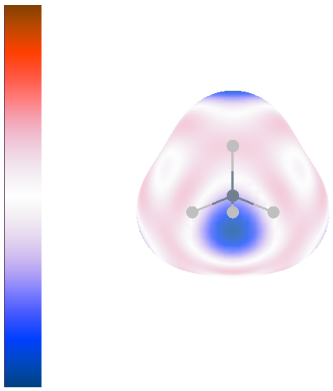


Figure 3.20: Differences in kJ mol$^{-1}$ between the rank 4 multipole model and charge-only model for methane at a surface density of $10^{-2}$ a.u. Maximum errors are given in Table 3.19. The range is $\pm 30$ kJ mol$^{-1}$, red indicates more positive charge for the charge-only model.

| | ISA charges |
|---|---|
| Max. Dif. | 9.899 |
| Min. Dif. | $-28.413$ |
| rmse | 8.524 |

Table 3.19: Differences in kJ mol$^{-1}$ between the rank 4 multipole model for methane and the ISA charges at a surface density of $10^{-2}$ a.u. The surface is shown in Fig 3.20. Positive difference indicates more positive charge for the charge-only model.

### 3.3.2 Polarisation Model

Unlike in the simplification process for the water model, there were no problems in including the undamped polarisability and so the polarisation model used the L1 isotropic components of the existing model.

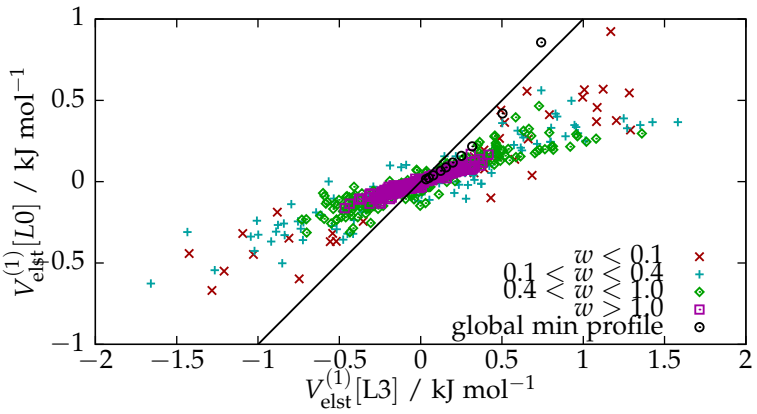



Figure 3.21: Scatter plot of the electrostatic energy using only the ISA charges against the electrostatic energy using the full distributed multipoles.

Later (Section 4.2.2), following the fit of the exchange repulsion terms, a basin-hopping search was carried out for a cluster of six methane molecules using the non-polarisable model and several structures were reminimised using the polarisable model. It was found that the structures and energies were almost identical with either model, see Fig 4.12 and Table 4.8. Since the effect of non-additivity was very small, it was concluded that polarisation was unnecessary for the methane model.

### 3.3.3 Short-Range Energy

The short-range energy was fitted twice, once with a polarisation model included and once without. Each fit used the terms from the anisotropic potential as anchor values and the same set of 2648 points; and applied a statistical weight $w = e^{-(E_{\text{int}}^{(1-\infty)} - E_{\text{min}})/E_0}$ to each point. When fitting the short-range energy it was assumed that the current set of data points would lead to a good overall fit, particularly as methane is often considered near-isotropic. However, the reference points along the profile of the global minimum, which were not included in the dataset, did not fit well and tended to underbind. In addition, the fitted potentials both had a global minimum around half a Bohr further out than the accepted result, suggesting that the preferred carbon-carbon separation along the global minimum profile is uncharacteristically low. On reflection this last point is not surprising: The global minimum of the methane dimer has the tetrahedral molecules arranged "face-to-face", see Chapter 4 for further discussion.

#### 3.3.3.1 Polarisable model

For the polarisable model, the anchor strengths of all repulsive parameters and $C_6$ dispersion coefficients were set and adjusted at the same value. $E_0$ was chosen first, keeping anchor strengths at 0.001. $E_0 = 2$ kJ mol$^{-1}$ was decided on as a compromise of fitting the global minimum and lowering the rmse for the lower energy categories, see Table 3.20 and Figures 3.22 and 3.23. Keeping $E_0 = 2$ kJ mol$^{-1}$, the anchor strengths were varied and it was found that $A = 0.01$ gave a better fit, see Table 3.21 and Figures 3.24 and 3.25. Setting $A = 0.1$ would give a better fit to the global minimum, however as previously mentioned this configuration is perhaps atypical, and so was not relied on too heavily as a guide for an isotropic fit. The final set of parameters for the potential, including its polarisabilty, is given in Table 3.22.

| $E_0$ kJ mol$^{-1}$ | 1 | 2 | 5 |
|---|---|---|---|
| $w > 1$ | 0.251 | 0.180 | 0.163 |
| $> 0.4, < 1$ | 1.433 | 0.649 | 0.475 |
| $> 0.1 < 0.4$ | 4.423 | 1.621 | 1.129 |
| $< 0.1$ | 11.872 | 4.668 | 4.495 |

Table 3.20: rmse in kJ mol$^{-1}$ from varying weighting parameter $E_0$ when fitting the polarisable isotropic methane model. These fits used anchor strength $A = 0.1$ and anchor values as shown in Table 2.15 for the exchange-repulsion parameters, and $C_6$ parameters as shown in Table 2.3. $w$ as in Table 2.13. $E_0 = 2$ kJ mol$^{-1}$ was chosen.

| A | 0.1 | 0.01 | 0.001 |
|---|---|---|---|
| $w > 1$ | 0.180 | 0.107 | 0.098 |
| $> 0.4, < 1$ | 0.649 | 0.444 | 0.431 |
| $> 0.1 < 0.4$ | 1.621 | 1.112 | 1.143 |
| $< 0.1$ | 4.668 | 4.128 | 4.304 |

Table 3.21: rmse in kJ mol$^{-1}$ from varying anchors strength $A$. These fits used weighting parameter $E_0 = 2$ kJ mol$^{-1}$ and anchor values as shown in Table 2.15 for the exchange-repulsion parameters, and $C_6$ parameters as shown in Table 2.3. $w$ as in Table 2.13. $A = 0.01$ was chosen.

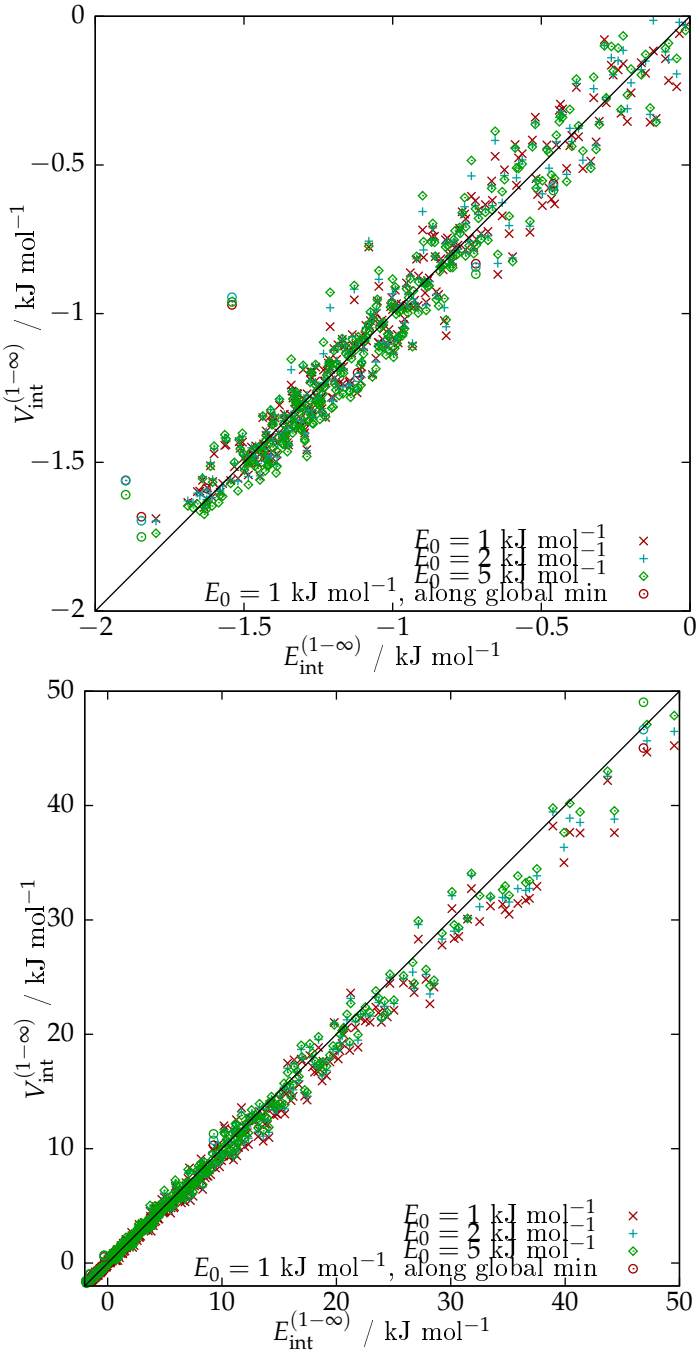



Figure 3.22: Scatter plot of $V_{\text{int}}^{(1-\infty)}$ against $E_{\text{int}}^{(1-\infty)}$ varying weighting parameter $E_0$ when fitting the polarisable isotropic methane model to the infinite-order energy with $A = 0.1$. Plots are over different energy ranges. Circular markers indicate points on the profile of the global minimum. $E_0 = 2$ kJ mol$^{-1}$ was chosen.

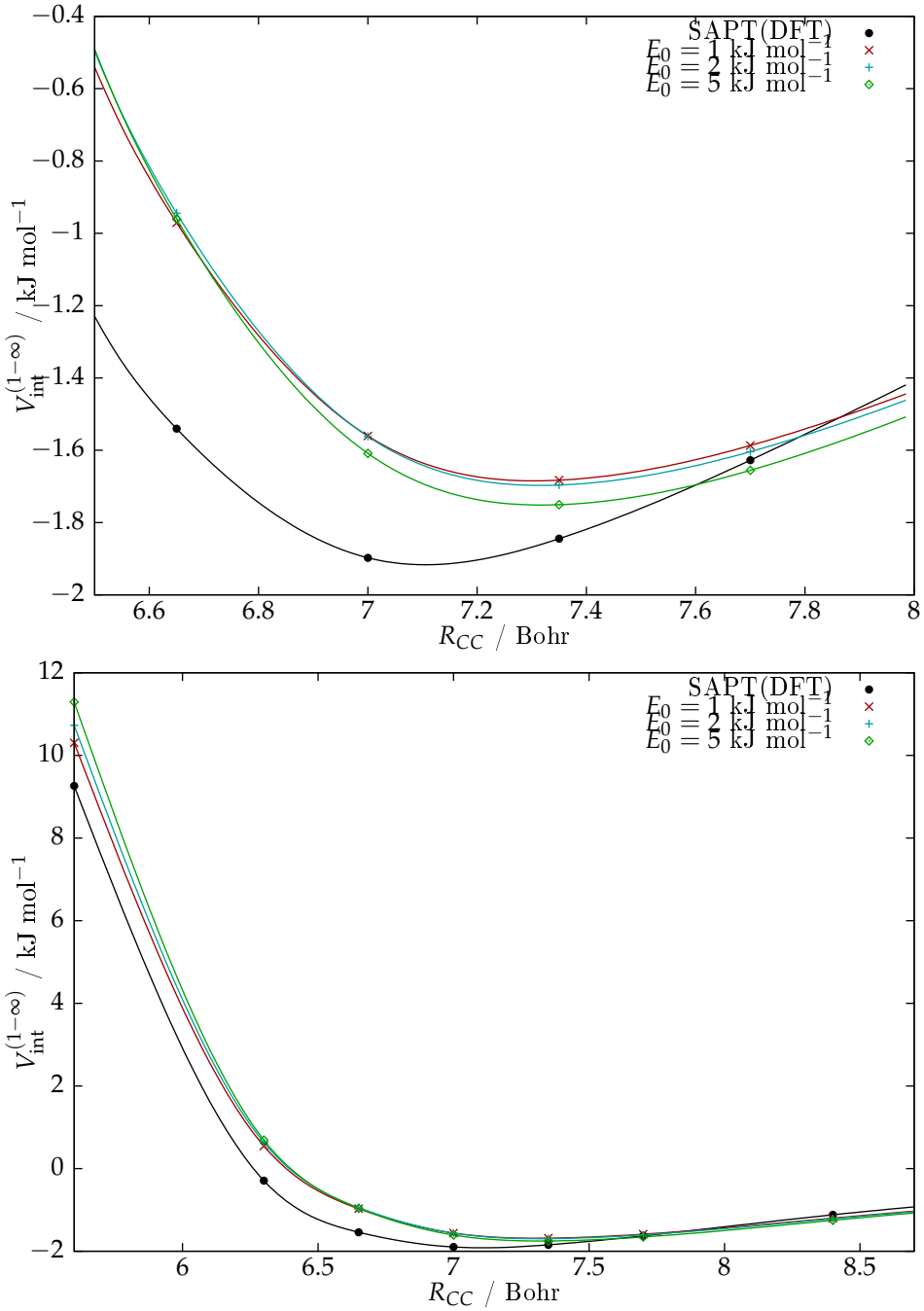



Figure 3.23: Profile plot of $V_{\text{int}}^{(1-\infty)}$ against carbon-carbon separation $R_{CC}$, varying weighting parameter $E_0$ when fitting the polarisable isotropic methane model to the infinite-order energy with $A = 0.1$. $E_0 = 2$ kJ mol$^{-1}$ was chosen.

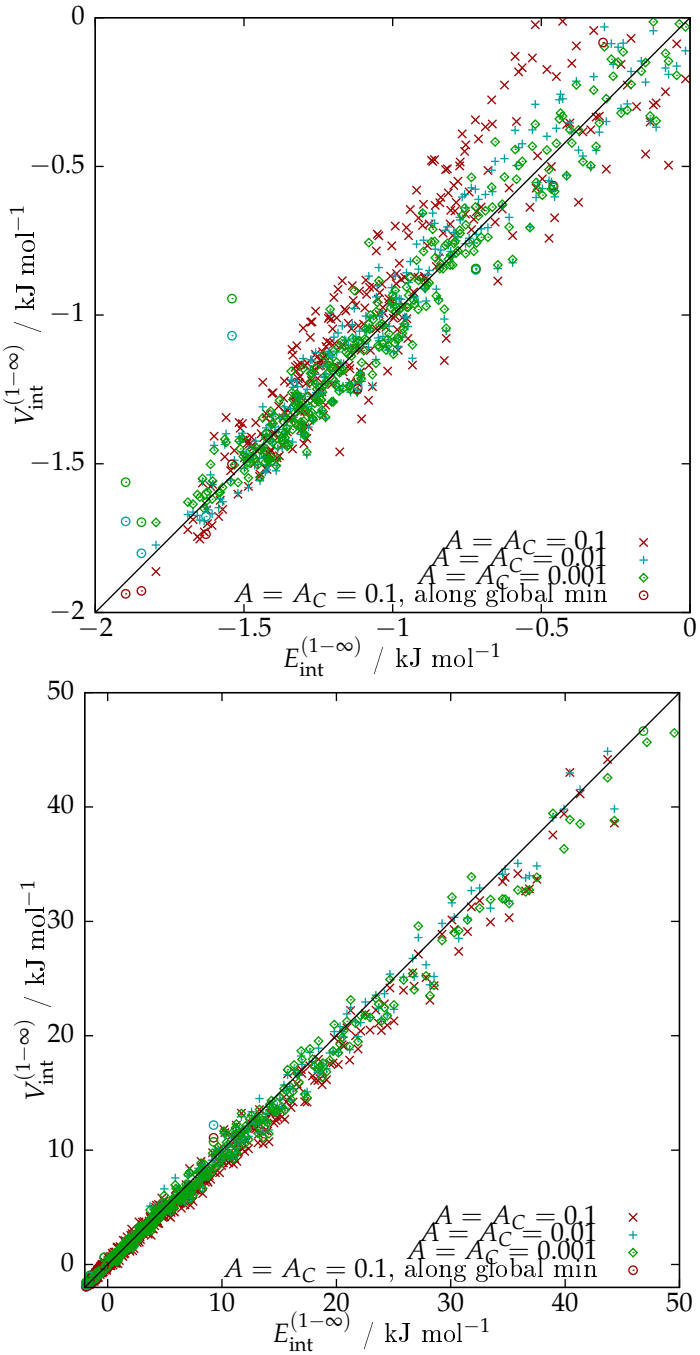



Figure 3.24: Scatter plot of $V_{\text{int}}^{(1-\infty)}$ against $E_{\text{int}}^{(1-\infty)}$ varying anchor strengths when fitting the polarisable isotropic methane model to the infinite-order energy with $E_0 = 2$ kJ mol$^{-1}$. Plots are over different energy ranges. Circular markers indicate points on the profile of the global minimum. $A = 0.001$ was chosen.

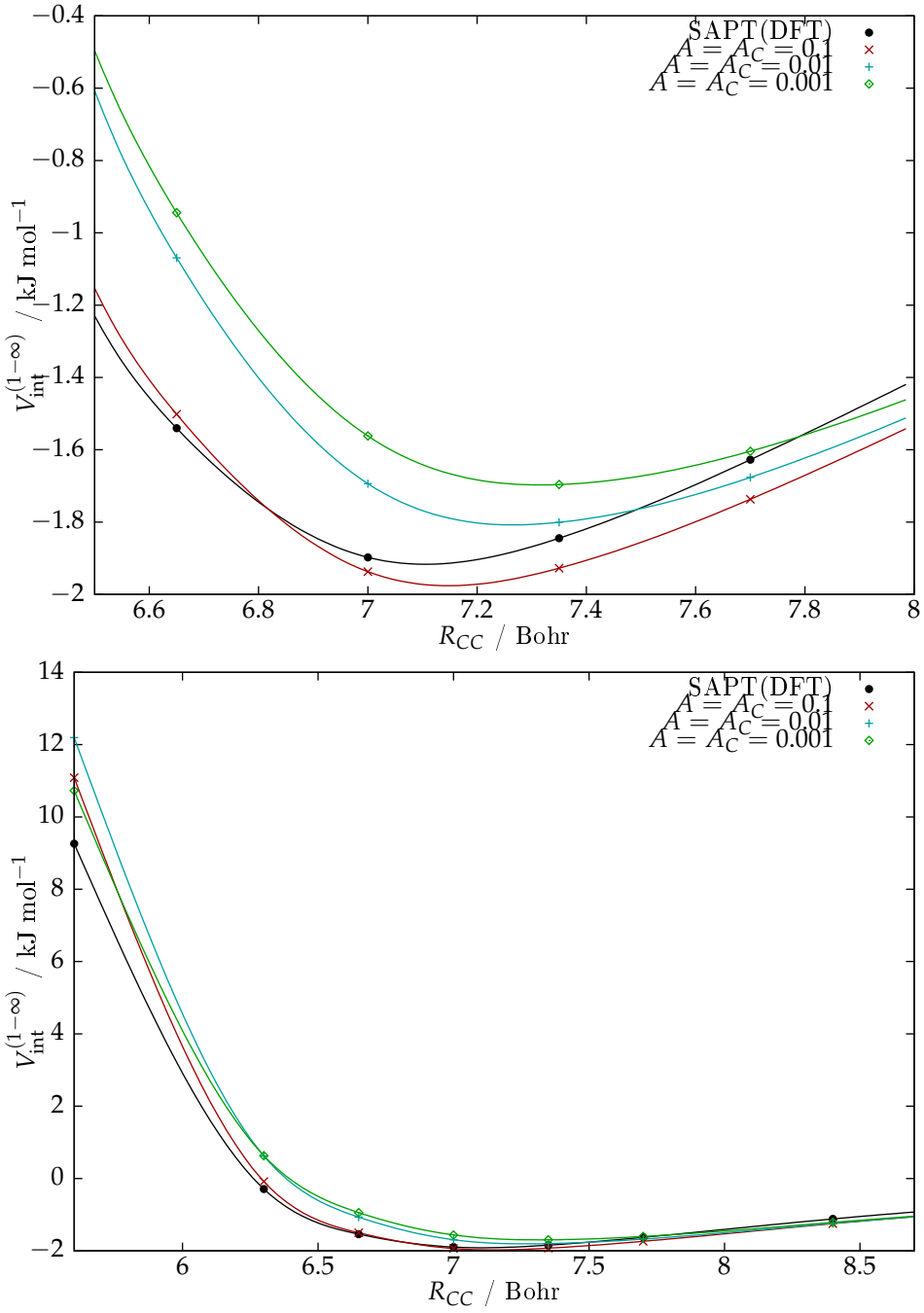



Figure 3.25: Profile plot of $V_{\text{int}}^{(1-\infty)}$ against carbon-carbon separation $R_{CC}$, varying anchor strengths when fitting the polarisable isotropic methane model to the infinite-order energy with $E_0 = 2$ kJ mol$^{-1}$. $A = 0.001$ was chosen.

| Shape functions | | | |
|---|---|---|---|
| Pair | $\alpha_{ab}$ | $\rho_{ab}$ | $C_6$ |
| C-C | 2.146 | 5.648 | 22.07923 |
| C-H | 1.867 | 4.228 | 10.09522 |
| H-H | 1.681 | 4.813 | 4.87260 |
| Polarisabilities | | | |
| C | 5.8092410 | | |
| H | 5.8092410 | | |

Table 3.22: Potential parameters for the polarisable isotropic methane-methane model.

#### 3.3.3.2 Non-polarisable model

For the non-polarisable model, $E_0 = 2$ kJ mol$^{-1}$ and $A = 0.01$ appeared good choices in an effort to balance the scatter and the fit to the global minimum. Allowing the $C_6$ coefficients to vary made very little difference to the fit, however slight improvement was observed for anchor strength $A_C = 0.001$. Although this is a very weak anchor, the dispersion coefficients changed very little, for example the C-C term went from 22.079 to 22.022 AU. The final set of parameters for the potential is given in Table 3.25.

| $E_0$ kJ mol$^{-1}$ | 1 | 2 | 5 |
|---|---|---|---|
| $w > 1$ | 0.251 | 0.181 | 0.164 |
| $> 0.4, < 1$ | 1.432 | 0.653 | 0.480 |
| $> 0.1 < 0.4$ | 4.411 | 1.627 | 1.143 |
| $< 0.1$ | 11.842 | 4.674 | 4.507 |

Table 3.23: rmse in kJ mol$^{-1}$ from varying weighting parameter $E_0$ for the non-polarisable methane model. These fits used anchor strength $A = 0.1$ and anchor values as shown in Table 2.15 for the exchange-repulsion parameters, and $C_6$ parameters as shown in Table 2.3. $w$ as in Table 2.13. $E_0 = 2$ kJ mol$^{-1}$ was chosen.

| A | 0.1 | 0.01 | 0.001 |
|---|---|---|---|
| $w > 1$ | 0.181 | 0.108 | 0.099 |
| $> 0.4, < 1$ | 0.653 | 0.449 | 0.436 |
| $> 0.1 < 0.4$ | 1.627 | 1.124 | 1.158 |
| $< 0.1$ | 4.674 | 4.137 | 4.320 |

Table 3.24: rmse in kJ mol$^{-1}$ from varying anchors strength $A$ for the non-polarisable methane model. These fits used weighting parameter $E_0 = 2$ kJ mol$^{-1}$ and anchor values as shown in Table 2.15 for the exchange-repulsion parameters, and $C_6$ parameters as shown in Table 2.3. $w$ as in Table 2.13. $A = 0.01$ was chosen.

| Pair | $\alpha_{ab}$ | $\rho_{ab}$ | $C_6$ |
|---|---|---|---|
| C-C | 1.821205 | 6.379491 | 22.022287 |
| C-H | 1.695672 | 4.825019 | 9.8579057 |
| H-H | 1.576686 | 4.302228 | 5.3231358 |

Table 3.25: Potential parameters in a.u. for the non-polarisable isotropic methane-methane model.

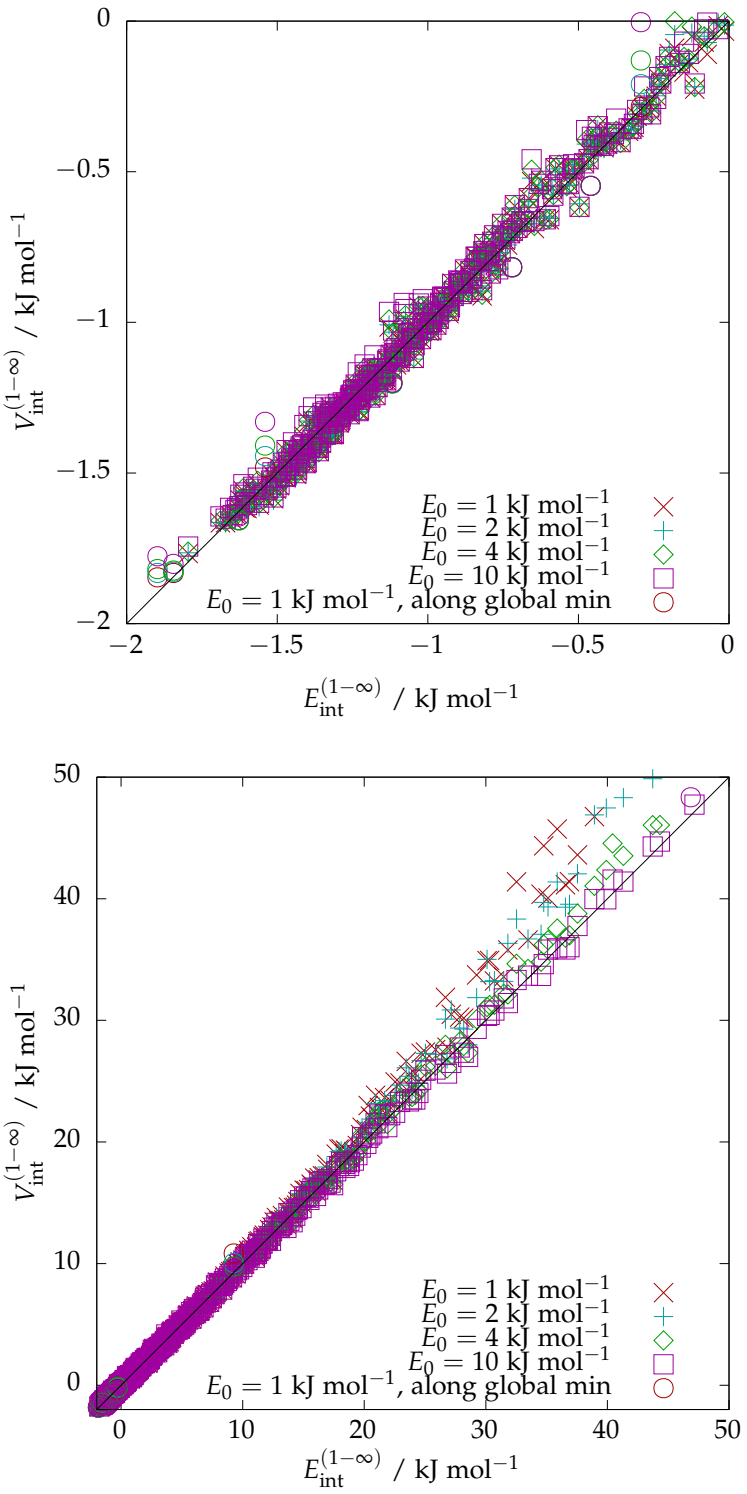



Figure 3.26: Scatter plot of $V_{\text{int}}^{(1-\infty)}$ against $E_{\text{int}}^{(1-\infty)}$ varying weighting parameter $E_0$ when fitting to the infinite-order energy with $A = 0.001$ for the non-polarisable methane model. Plots are over different energy ranges. Circular markers indicate points on the profile of the global minimum. $E_0 = 2$ kJ mol$^{-1}$ was chosen.

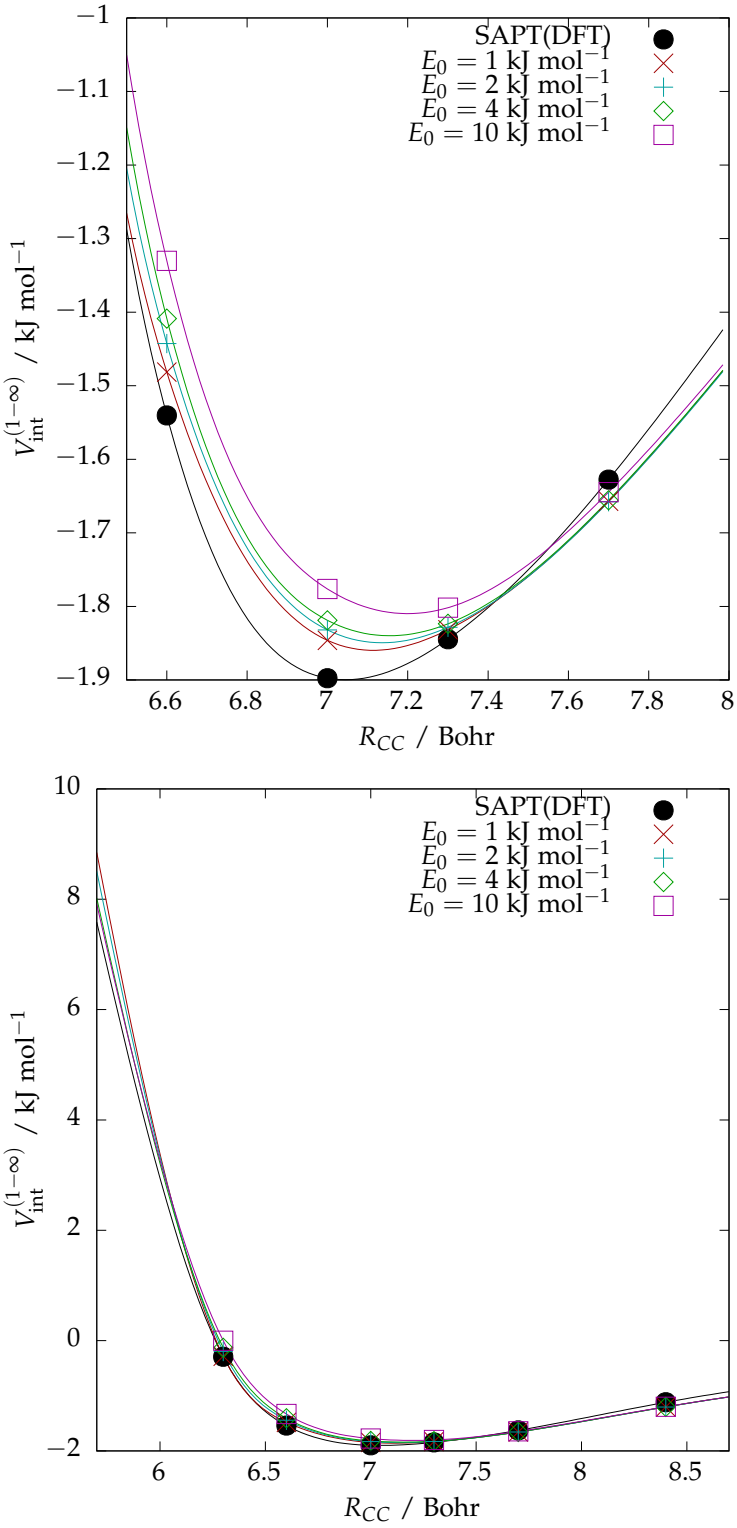



Figure 3.27: Profile plot of $V_{\text{int}}^{(1-\infty)}$ against carbon-carbon separation $R_{CC}$, varying weighting parameter $E_0$ when fitting to the infinite-order energy with $A = 0.001$ for the non-polarisable methane model. $E_0 = 2$ kJ mol$^{-1}$ was chosen.

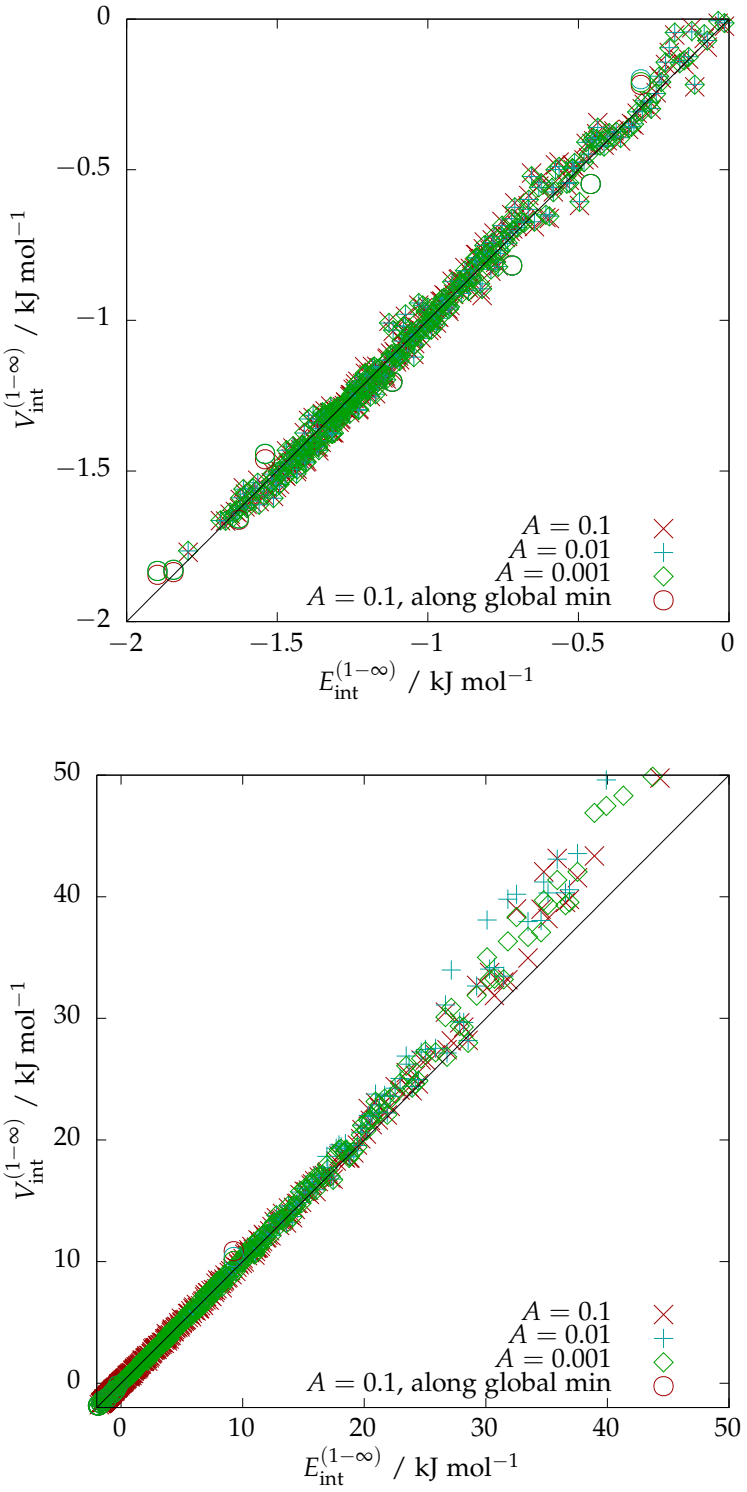



Figure 3.28: Scatter plot of $V_{\text{int}}^{(1-\infty)}$ against $E_{\text{int}}^{(1-\infty)}$ varying weighting parameter $E_0$ when fitting to the infinite-order energy with $A = 0.001$ for the non-polarisable methane model. Plots are over different energy ranges. Circular markers indicate points on the profile of the global minimum. $E_0 = 2$ kJ mol$^{-1}$ was chosen.

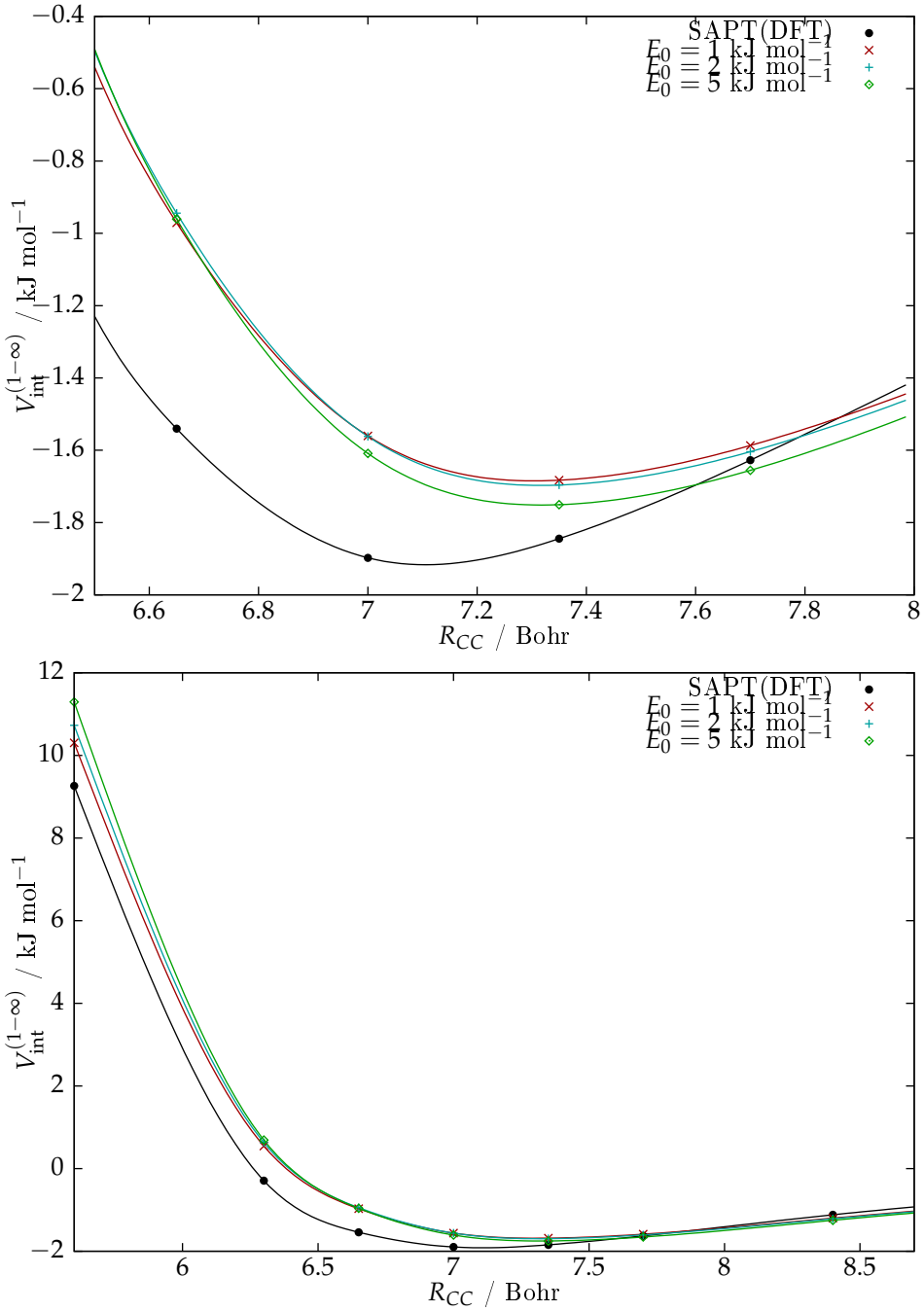



Figure 3.29: Profile plot of $V_{\text{int}}^{(1-\infty)}$ against carbon-carbon separation $R_{CC}$, varying anchor strengths when fitting to the infinite-order energy with $A = 0.001$ for the non-polarisable methane model. $E_0 = 2$ kJ mol$^{-1}$ was chosen.

### 3.3.4 Methane potential with Lennard-Jones interactions

The long-standing popularity of Lennard-Jones potentials led to the development of a methane potential with Lennard-Jones sites on each atom. As the model in Section 3.3.3.2 consists of a charge model, a repulsive wall and a dispersion term of the form $-N/r^6$; loosely speaking we remove the dispersion damping and replace the repulsive wall, which takes an exponential form, with a term of the form $M/r^{12}$. There was little success attempting to match both the global minimum and the other points, this was even more the case than in the previous section.

#### 3.3.4.1 Fitting parameters

For intitial guesses at the terms $A$ and $B$, where $E_{\text{int}}^{(1-\infty)} = M/r^{12} - N/r^6$, the $C_6$ term was used as $N$ and $M$ was then chosen so as to give the same value as the potential in Table 2.20 at $r = 7$ Bohr for the C-C term and at $r = 5$ Bohr for the C-H and H-H terms, these separations corresponding very roughly to those found in the global minimum. These parameters were used as starting values for terms fitted using Orient, but the term $M$ was not anchored during the fit (i.e. anchor strength of zero). While fitting the anchors $E_0$ was set to 2 kJ mol$^{-1}$, based on the minimum depth and on fitting the previous methane models.

Reducing the anchor strength $A$ led to a reduced rmse in all energy categories, however it also led a poorer fit the minimum. In order to keep any distinguished minimum at all the anchors were kept at $A = 0.1$. Reducing the weighting parameter to $E_0 = 1$ kJ mol$^{-1}$led to a slight improvement in well depth and a slight reduction in rmse for negative energies (0.449 kJ mol$^{-1}$compared with 0.564 kJ mol$^{-1}$). Although it also led to much higher rmse in other categories as shown in Table 3.26, if methane is taken to be close to isotropic it can be argued that any two interacting methane molecules in bulk methane are able to orient themselves to a low energy. This is taken as the potential, with parameters given in Table 3.27.

| $E_0 = 2$ kJ mol$^{-1}$, $A =$ | 0.1 | 0.01 | 0.001 | $E_0 = 1$ kJ mol$^{-1}$, $A = 0.1$ |
|---|---|---|---|---|
| $w > 1$ | 0.564 | 0.298 | 0.281 | 0.449 |
| $> 0.4, < 1$ | 3.501 | 1.954 | 1.541 | 8.218 |
| $> 0.1 < 0.4$ | 28.657 | 18.557 | 14.942 | 9.146 |
| $< 0.1$ | 96.229 | 62.526 | 52.544 | 151.145 |

Table 3.26: rmse in kJ mol$^{-1}$ from varying $A$, the anchor strength of Lennard-Jones parameter $N$; and weighting parameter $E_0$, when fitting the Lennard-Jones methane model. The anchor values are as shown in Table 3.27. $w$ as in Table 2.13. $E_0 = 1$ kJ mol$^{-1}$, $A = 0.1$ was chosen.

| Pair | Initial $M$ | Fitted $M$ | Initial $N$ | Fitted $N$ |
|---|---|---|---|---|
| C-C | 4470860 | 10539773 | 22.022 | 22.003 |
| C-H | 181460 | 440330.68 | 9.858 | 9.629 |
| H-H | 81254 | 25139.086 | 5.323 | 4.545 |

Table 3.27: Potential parameters $M$ and $N$ in the Lennard-Jones methane-methane fit, given in atomic units.

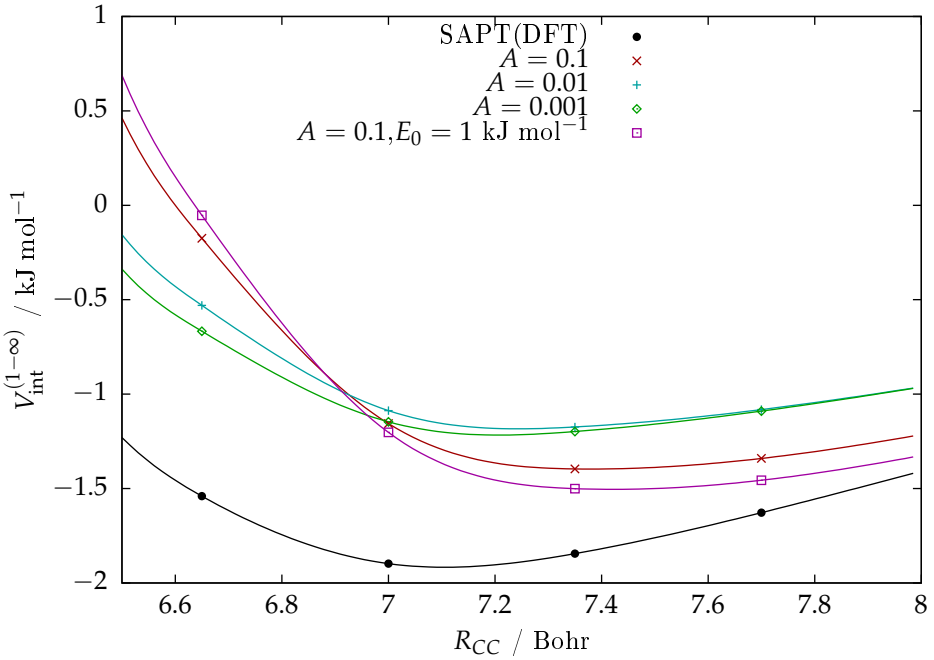



Figure 3.30: Profile plot of total interaction energy along the profile of the global minimum methane dimer varying weighting parameter $E_0$ and anchor strength $A$ for the term $N/r^6$, when fitting the Lennard-Jones methane model. Anchor values as shown in Table 3.27. $A = 0.1, E_0 = 1$ kJ mol$^{-1}$ was chosen.

## 3.4 Simplifying the Water-Methane Potential

In simplifying the water-methane pair potential only the repulsive wall terms had be decided. For the initial guesses, the values of $\alpha$ were taken directly from the anisotropic water-methane potential; whereas the values of $\rho$ were taken from the isotropic water-water and methane-methane pair potentials (where the SAPT-5s charge locations are used and the methane is non-polarisable).

In order to come as close as possible to the correct depth at the global minimum, the anchor strength was set to $A = 0.001$ and the weighting parameter was set to $E_0 = 1$ kJ mol$^{-1}$, see Figures 3.31 and 3.32. The adjustment of$A$ had little effect on rmse, but setting $E_0 = 1$ kJ mol$^{-1}$did give a higher rmse for low positive energies than would have been the case if a value of 2 kJ mol$^{-1}$ was used instead, see Table 3.28.

Both the input and output parameters are given in Table 3.30.

| $E_0$ / kJ mol$^{-1}$ | 1 | 2 | 5 |
|---|---|---|---|
| $w > 1$ | 0.473 | 0.444 | 0.483 |
| $0.4 < w < 1$ | 2.426 | 2.049 | 1.750 |
| $0.1 < w < 0.4$ | 13.176 | 9.231 | 5.605 |
| $w < 0.1$ | 48.699 | 32.028 | 14.648 |

Table 3.28: rmse in kJ mol$^{-1}$ varying the weighting parameter $E_0$ when fitting the exchange-repulsion terms for the isotropic water-methane model, keeping anchor strength $A = 0.1$. $E_0 = 1$ kJ mol$^{-1}$ was chosen.

| $A$ | 0.1 | 0.01 | 0.001 |
|---|---|---|---|
| $w > 1$ | 0.473 | 0.442 | 0.435 |
| $0.4 < w < 1$ | 2.426 | 2.668 | 2.739 |
| $0.1 < w < 0.4$ | 13.176 | 14.446 | 14.943 |
| $w < 0.1$ | 48.699 | 49.901 | 50.497 |

Table 3.29: rmse in kJ mol$^{-1}$ varying the anchor strength $A$ when fitting the exchange-repulsion terms for the isotropic water-methane model, keeping the weighting parameter $E_0 = 1$ kJ mol$^{-1}$. $A = 0.001$ was chosen.

| Pair | $\alpha_{ab}$ (in) | $\alpha_{ab}$ (out) | $\rho_{ab}$ (in) | $\rho_{ab}$ (out) |
|---|---|---|---|---|
| C-O | 2.106968 | 2.044257 | 6.196215 | 6.338081 |
| C-Hw | 1.539480 | 2.708359 | 5.3468545 | 4.026565 |
| H-O | 1.756528 | 2.143781 | 5.1575835 | 4.550381 |
| H-Hw | 1.890521 | 1.710300 | 4.308223 | 3.814876 |

Table 3.30: Short-range input and output potential parameters in a.u. for the isotropic water-methane model.

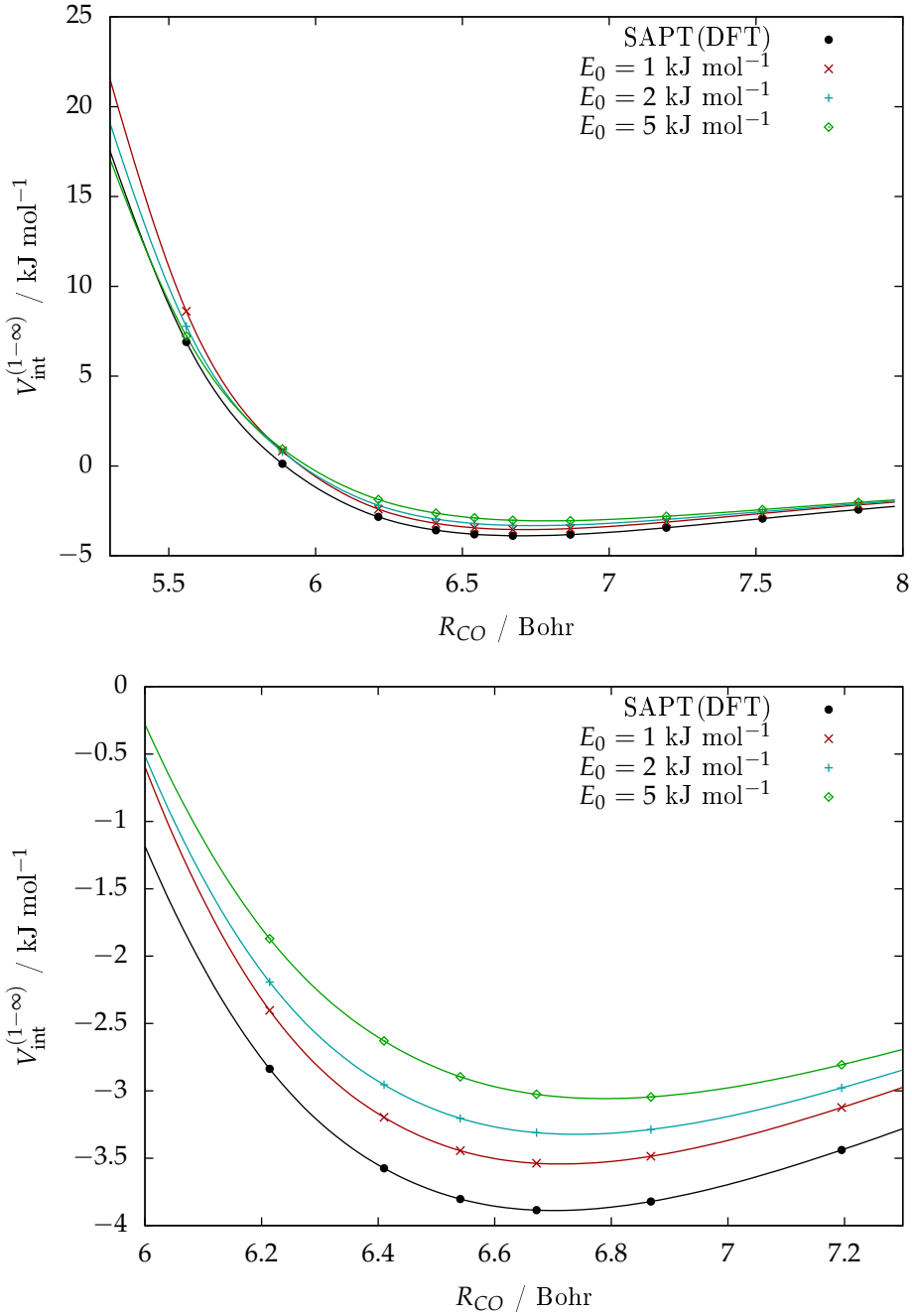



Figure 3.31: Profile plot of $V_{\text{int}}^{(1-\infty)}$ against carbon-oxygen separation $R_{CO}$, varying weighting parameter $E_0$ with anchor strength $A = 0.1$. $E_0 = 1$ kJ mol$^{-1}$ was chosen.

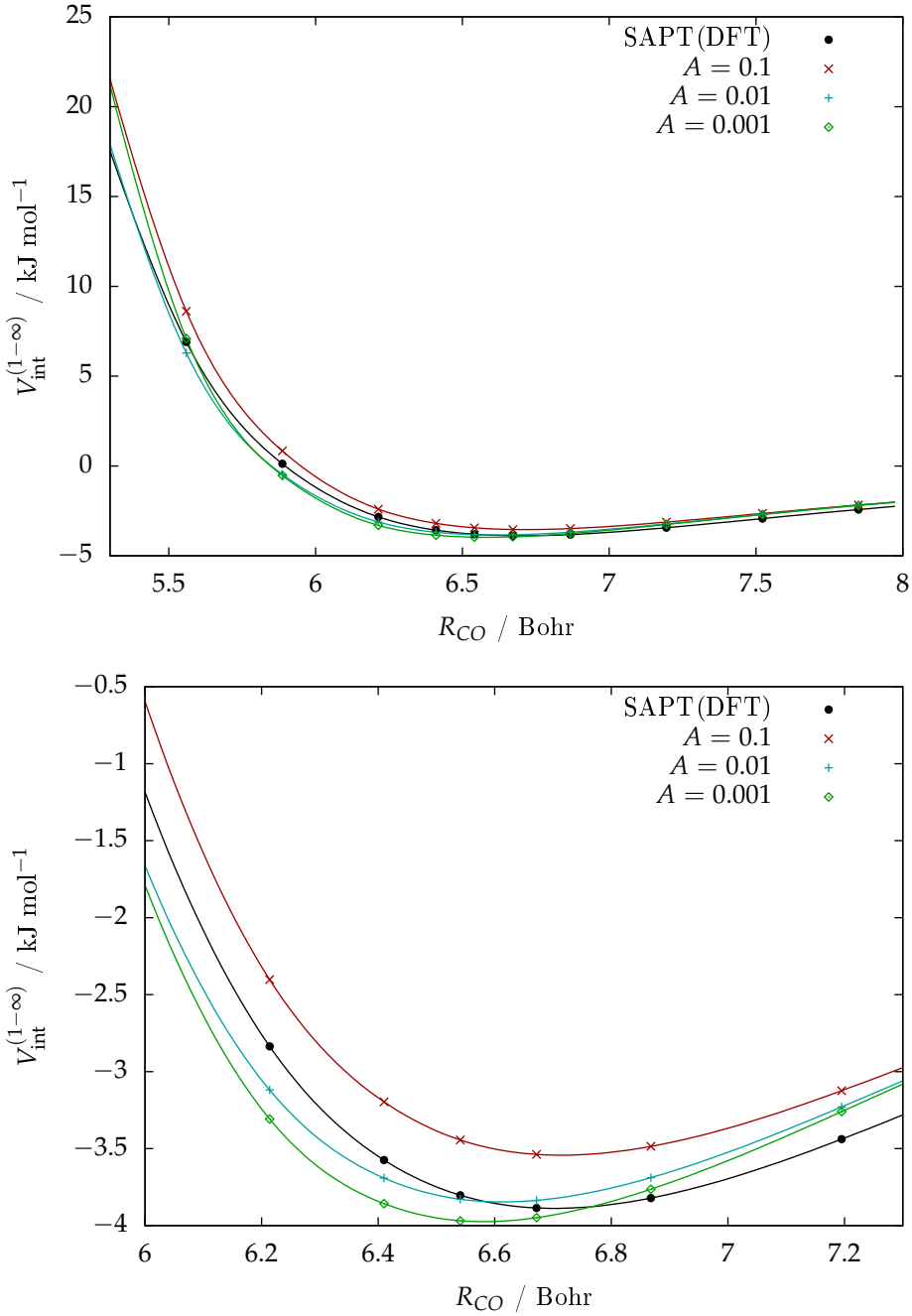



Figure 3.32: Profile plot of $V_{\text{int}}^{(1-\infty)}$ against carbon-oxygen separation $R_{CO}$, varying anchor strength $A$ with weighting parameter $E_0 = 1$ kJ mol$^{-1}$. $A = 0.001$ kJ mol$^{-1}$ was chosen.

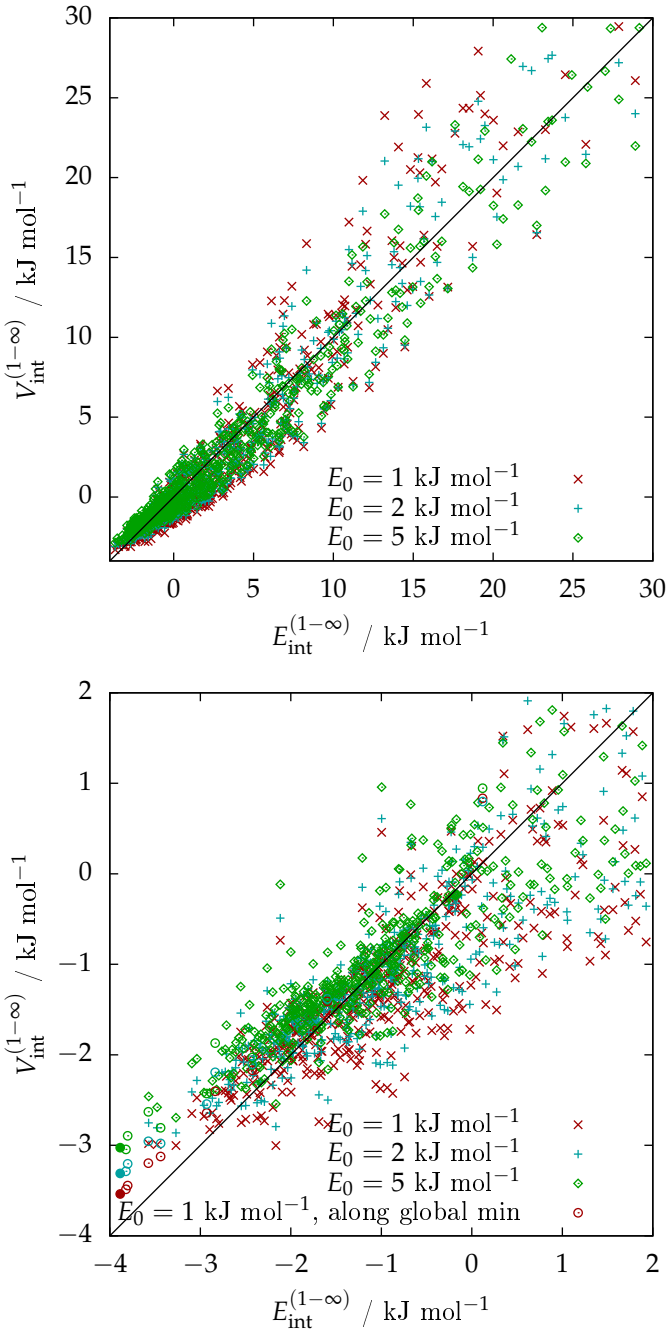



Figure 3.33: Scatter plot of $V_{\mathrm{int}}^{(1-\infty)}$ against $E_{\mathrm{int}}^{(1-\infty)}$, varying weighting parameter $E_0$ with anchor strength $A = 0.1$ when fitting the isotropic water-methane model. Plots are over different energy ranges. Circular markers indicate points on the profile of the global minimum. $E_0 = 1$ kJ mol$^{-1}$ was chosen.

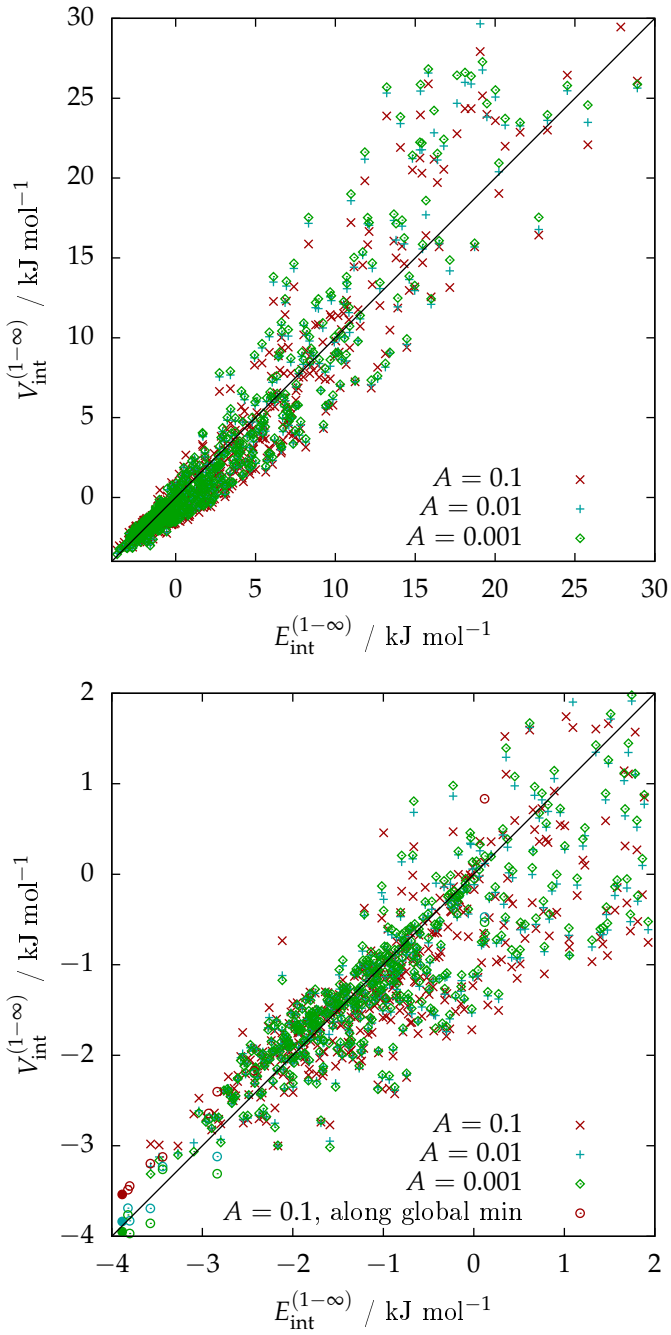



Figure 3.34: Scatter plot of $V_{\text{int}}^{(1-\infty)}$ against $E_{\text{int}}^{(1-\infty)}$, varying anchor strength $A$ with weighting parameter $E_0 = 1$ kJ mol$^{-1}$when fitting the isotropic water-methane model. Plots are over different energy ranges. Circular markers indicate points on the profile of the global minimum. $A = 0.001$ kJ mol$^{-1}$ was chosen.

## 3.5 Summary

The potentials from Chapter 2 have been used as a start point to produce much-simplified potentials with the intent that they can be used in MD simulations with DL_POLY4, with only the dispersion models remaining fully intact.

For methane, the transition method was relatively straight-forward and it was found polarisability could be removed entirely; but it became clear that an anisotropic charge-only potential would not be able to capture the depth or well-position of the global minimum if it was expected to describe other geometries with any reasonable accuracy. In addition to the model based closely on the previous chapter, a model using Lennard-Jones sites for each atom was used. This model performed worse both at describing the minimum and fitting the reference points.

For water, several electrostatic models using additional charge sites were used. The lack of damping available for polarisation meant that a seven-site charge model had to be discarded, and the best alternative was to use the charge locations of the SAPT-5s potential.

The water short-range potential was fitted with the understanding that rank 1 undamped point-polarisabilities on each atom site could be satisfactorily approximated on DL_POLY. This was not the case, and several possible methods to account for the polarisation model were discussed, including a shell model allowing for explicit polarisation and methods of altering the charge models to make each molecule behave as if it were polarised.

The performance of the models developed here is discussed further in Chapter 4.

# Chapter 4

# Validation of Models

Tests of the models developed can be split into two categories:

Properties of a small number of molecules:

- The potential along the global minimum profile
- Energies and structures of the dimer and clusters
- Second virial coefficient

Bulk properties:

- Radial distribution function
- Diffusion coefficient
- Melting point

Few-body properties are found using Orient, whereas bulk properties are found using MD simulation using DL_POLY. Because of this, not all tests can be performed on all models. In the majority of cases it is possible to compare results with benchmark data, however no other data is available on the methane clusters and water-methane clusters looked at, so these serve only at as a way to compare the potentials developed in the previous two chapters at each stage of simplification.

## 4.1 The Global Minimum

### 4.1.1 Water Dimer

The minimum-energy configuration of the water dimer contains one plane of symmetry; in the plane of the hydrogen-donating molecule; and can be parameterised by the O-O separation $R_{OO}$, the angle between the donor hydrogen, the oxygen of its water molecule and the other oxygen, and the angle between the O-O line and the line of symmetry of the molecule whose plane is perpendicular to it; as shown in Fig 4.1.

The minimum for the anisotropic potential was found using a basin-hopping search which used 500 steps and found two configurations resembling the correct minimum within 0.01 kJ $mol^{-1}$ of one another. The minimum configurations for the isotropic potentials were found by reminimising the lower-energy of these two geometries.

Measurements for our models and others are shown in Table 4.1. For all parameters, the anisotropic model performs best, followed by the CC-Pol model of Bukowski et al. (2008) [130]. Also for all parameters, our isotropic polarisable model performs better than TIP4P/2005 and TIP4P/Ice, but when this is simplified to a non-polarisable model the angle measurements become poorer than either TIP4P model.

| | $R_{OO}$/ Å | $\Theta/^\circ$ | $\Phi/^\circ$ | $E_{int}$/ kJ mol$^{-1}$ |
|---|---|---|---|---|
| Aniso | 2.917 | 5.26 | 126.33 | −21.013 |
| iso | 2.897 | 1.64 | 133.02 | −20.567 |
| nonpol, $+\Delta Q$ | 2.863 | 0.47 | 144.96 | −29.259 |
| nonpol, $+\Delta Q/2$ | 2.93 | 0.48 | 117.93 | −22.653 |
| CC-Pol[a] | 2.911 | 6.48 | 122.66 | −22.179 |
| TIP4P/2005 | 2.77 | 1.25 | 130.98 | −28.721 |
| TIP4P/Ice | 2.784 | 1.12 | 131.83 | −31.745 |
| SAPT[b] | 2.953 | 4.6 | 124 | −20.334 |
| CCSD(T)[c] | 2.909 | 4.47 | 124.92 | −21.087 |

Table 4.1: Geometric parameters as shown in Fig 4.1 and total energy of the minimum-energy dimer from our models and other calculations. *a* From [130]. *b* From [14]. *c* Geometry optimised in TZ2P(f,d)+dif basis set, energy from MP2 CBS extrapolation, from [131].

### 4.1.2 Methane Dimer

The minimum-energy methane dimer is as shown in Fig 4.2 and has only one geometric parameter; chosen as the C-C separation $R_{CC}$. A basin-hopping search using the final potential from Section 2.6 gave a value of $R_{CC}$ within 0.15 Å of the separation found by Li and Chao (2006) [132]; and an energy within 0.12 kJ mol$^{-1}$ of the energy given in the same paper computed by extrapolating with the MP2 method. The isotropic, non-polarisable model has a separation less than 0.06 Å higher and an energy only 0.060 kJ mol$^{-1}$lower.

The Lennard-Jones potential, which was fitted to give separation close to the SAPT(DFT) value, has an energy 0.456 kJ mol$^{-1}$ higher than the MP2 reference value for this configuration. In fact, the correct global minimum is not the lowest-energy minimum using this potential by some stretch: The configuration shown in Fig 4.2(c) has an energy of -1.928 kJ mol$^{-1}$ compared with -1.510 kJ mol$^{-1}$ for this configuration.

Although methane is often regarded as close to isotropic, the minimum well seems to be a somewhat unusual profile that does not simply "fall out" when fitting to a random set of geometries. To illustrate this, profile plots are given for the first and second minima in Fig 4.3, with the latter much more closely matching SAPT(DFT) calculations.

Note that the global minimum has a carbon-carbon separation between 7.0 and 7.1 Bohr whereas the secondary minimum has a higher separation of around 7.8 Bohr. Additionally, the minimum found using the Lennard Jones potential also has a larger separation at 7.4 Bohr. There are obvious difficulties associated with finding the correct minimum on a broad, shallow potential, but the global minimum may be particularly difficult to fit due to the unusually short separation. Only a few minima have been observed, but it is reasonable to assume the minimum geometry encourages the closest C-C approach, since that arrangement of molecules maximises the distance between hydrogen atoms for a given C-C separation.

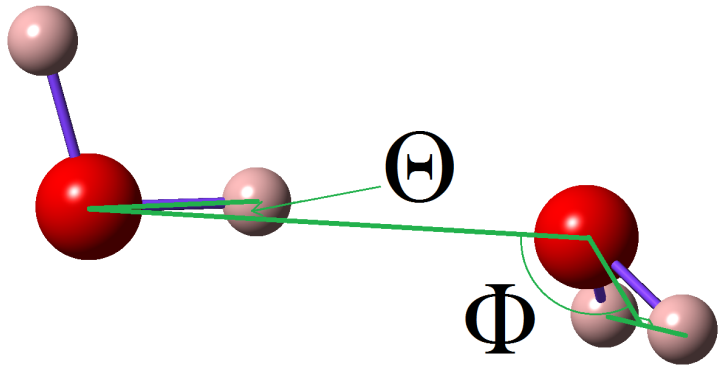



Figure 4.1: The minimum-energy configuration water dimer. The O-O separation and measurements for the angles indicated are given in in Table 4.1

.

| | Aniso | Iso-nonpol | LJ | MP2[a] |
|---|---|---|---|---|
| $R_{CC}$ Å$^{-1}$ | 3.803 | 3.857 | 3.938 | 3.66 |
| $E_{\text{int}}^{(1-\infty)}$ / kJ mol$^{-1}$ | −1.847 | −1.787 | −1.510 | −1.966 |

Table 4.2: Geometric parameter $R_{CC}$ as shown in Fig 4.2 and total energy $E_{\text{int}}^{(1-\infty)}$ of the minimum-energy dimer from our models and MP2 calculations. *a* From [132] at CBS limit.

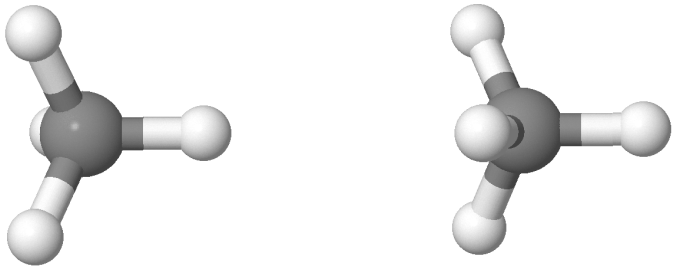
### 4.1.3 Methane-Water Dimer

The minimum-energy configuration for the methane-water dimer is shown in Fig 4.4 using either the anisotropic water and methane models or the isotropic non-polarisable methane model and a specified water model. Describing the dimer required three measurements, which we choose as separation $R_{OC}$, the angle $O-Hw-C$ and the angle $O-C-H$ as indicated in the figure. Literature often includes only the first of these two measurements, which is somewhat understandable given that; firstly, $O-C-H$ is close to 180°; and secondly that methane is besides usually considered to be close to isotropic. The anisotropic water model gives a minimum depth of only -4.00 kJ mol$^{-1}$, reference results give a minimum 0.31 kJ mol$^{-1}$deeper. The isotropic and $+\Delta Q$ models are within 0.1 kJ mol$^{-1}$of the anisotropic model, the $+\Delta Q/2$ model has a depth 0.26 kJ mol$^{-1}$ above the anisotropic model. All isotropic models overestimate the $O-C-H$ angle by about half a degree and the $O-Hw-C$ by about 7°.

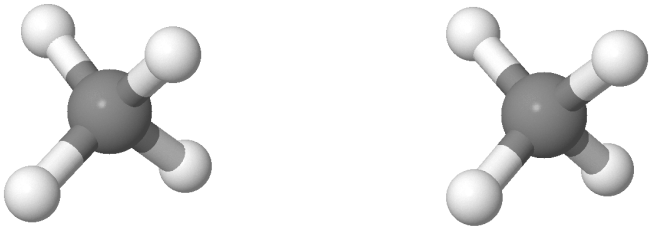
| | Aniso | Iso | nonpol water $+\Delta Q$ | nonpol water $+\Delta Q/2$ | CCSD(T)[a] |
|---|---|---|---|---|---|
| $R_{CO}$ Å$^{-1}$ | 3.530 | 3.525 | 3.499 | 3.533 | 3.51 |
| $O-C-H$ (°) | 173.40 | 177.31 | 178.28 | 177.80 | - |
| $O-Hw-C$ (°) | 165.63 | 172.23 | 173.34 | 172.76 | 165.6 |
| $E_{\text{int}}^{(1-\infty)}$ / kJ mol$^{-1}$ | −4.00 | −3.988 | −3.928 | −3.739 | −4.314 |

Table 4.3: Geometric parameters as shown in Fig 4.4 and total energy $E_{\text{int}}^{(1-\infty)}$ of the minimum-energy dimer from our models alongside CCSD(T) calculations. *a* From [40].

(a)



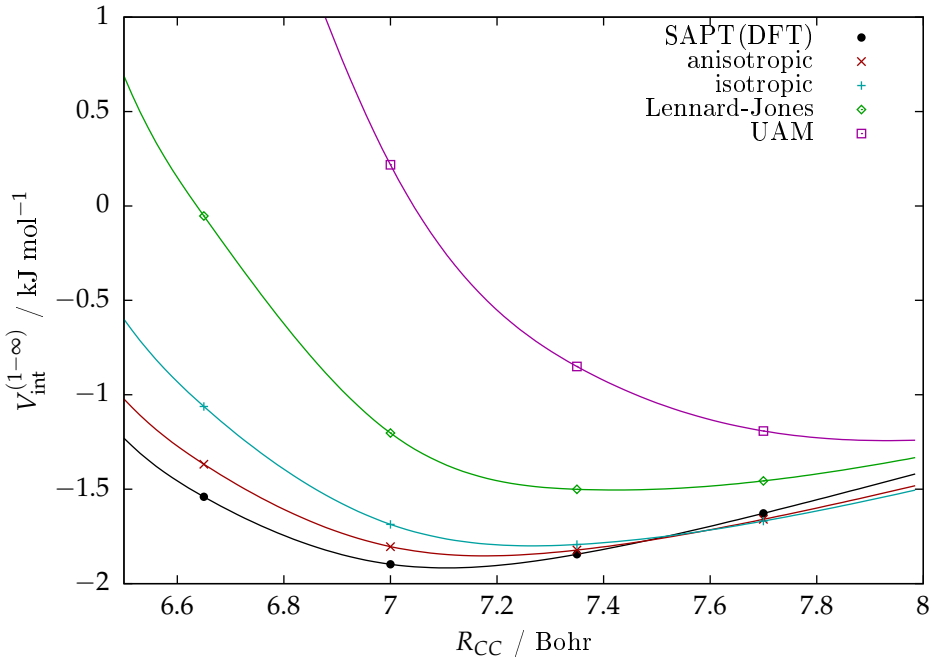
(b)



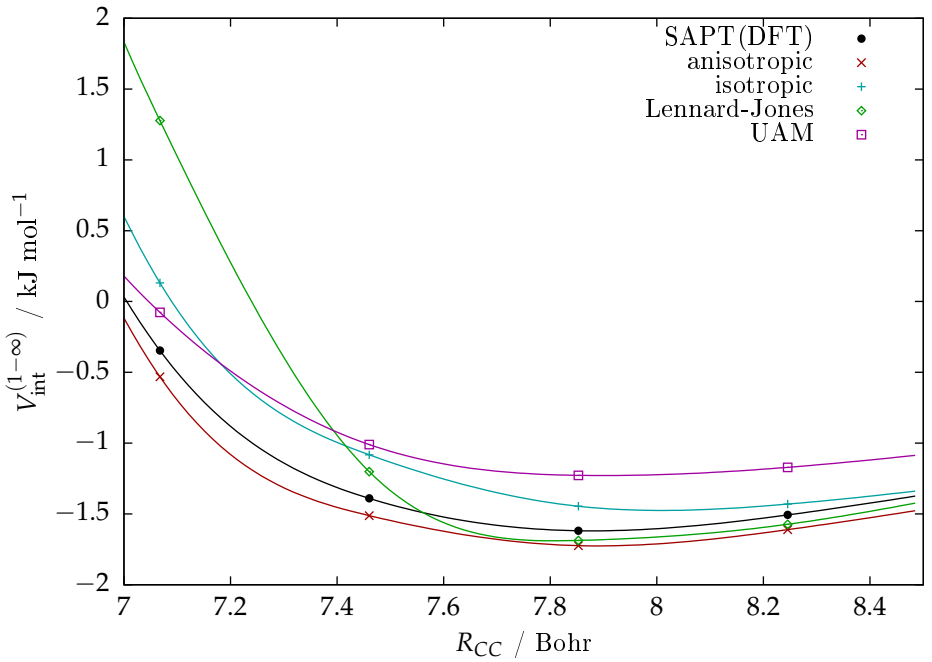
(c)

Figure 4.2: 4.2(a) The minimum-energy configuration methane dimer, 4.2(b) the secondary minimum, 4.2(c) the minimum found using the Lennard-Jones potential. Each dimer is parameterised only by the intermolecular separation, taken to be $R_{CC}$

.




(a)




(b)

Figure 4.3: Profile plot of total interaction energy along the profile of the global minimum(4.3(a)) and the secondary minimum(4.3(b)) shown for the three methane models developed here: The anisotropic and polarisable model, the isotropic nonpolarisable model, and the Lennard-Jones model. The single-site United Atom Methane model[35], which will later be used in some simulations, is also shown.

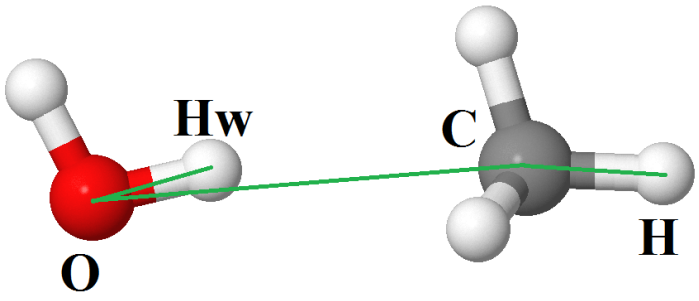



Figure 4.4: The minimum-energy configuration for the water-methane dimer. Measurements taken were separation $R_{OC}$, the angle $O - Hw - C$ and the angle $O - C - H$.

## 4.2 Energies and Structures of Clusters

The study of water clusters is a common topic when describing water. The hexamer structures are of particular interest because there are several structures within a few kJ $mol^{-1}$of one another, and because this is the smallest water cluster for which the lowest energy structure is not a ring.

The study of small clusters of methane and of water-methane is much less common, however given that one of the aims of developing these potentials is to simulate methane clathrate, information on water-methane clusters might be relevant.

At present it is not possible to test the anisotropic potentials in many-body simulations with available MD software. Consequently small clusters are the only way to test many-properties for the anisotropic models, and to compare the isotropic models. The computation times can also be compared; reminimising each hexamer for the anisotropic water model takes around 1 minute, for the polarisable isotropic model it takes around 10 seconds, and for non-polarisable isotropic models each reminimisation takes around half a second. For energy calculations, the set of 2048 water dimers takes on the order of 10 seconds for the anisotropic model, and less than a second for either isotropic model. For SAPT(DFT), each dimer energy calculation took around 3 minutes.

### 4.2.1 Water Clusters

The data given here includes the $R_{OO}$ separation between neighbours and the binding energies for the dimer and rings up to the hexamer ring, and the binding energies for five low-lying hexamers.

#### 4.2.1.1 Water rings

For the dimer separation the non-polarisable model still gives a closer agreement (0.06 Å) with our anisotropic model than either TIP4P/2005 or TIP4P/Ice (0.15 and 0.14 Å respectively). As the ring size increases, the average separation between nearest neighbours is expected to decrease, for our anisotropic model this decrease is from 2.80 Å for the trimer to 2.71 Å for the hexamer. This gradual decrease is about half that size for the TIP4P models (2.78 to 2.74 Å and 2.80 to 2.76 Å) and is not shown at all for our own nonpolarisable model (2.85 Å to 2.84 Å). In the isotropic polarisable model the change is actually larger; 2.86 Å to 2.68 Å. See Table 4.4 and Fig 4.5.

The energy of each ring can also be plotted for each model and compared with MP2 data at the CBS limit from Yoo et al. [133], see Fig 4.7 or Table 4.5. For all points the anisotropic model matches the MPS/CBS data most closely. The anisotropic and isotropic polarisable models both have a notable disagreement with the MP2/CBS data at the three-member ring (+4.4% and +16.4% respectively). In fact for the three-member ring these models prefer the wrong minimum: The trimer ring should have two hydrogens pointing "upwards" out of the plane and one pointing "downwards", whereas these two models give the trimer with all three hydrogens pointing "upwards", see Fig 4.6. This structure is estimated by Anderson et al. [134] to have an energy 3.21 kJ $mol^{-1}$ higher than the true global minimum. It could be argued that of all rings, the trimer least closely resembles the dimer, since it has the sharpest $O-O-O$

angle, this may explain why potentials fitted to match dimers have performed relatively poorly here. The non-polarisable model using $Q + \Delta Q$ agrees with the TIP4P/2005 model within 3 kJ mol$^{-1}$ for $n = 2 - 5$, then deviates by +7.8 kJ mol$^{-1}$ for the hexamer ring, bringing it close to (within 2.6 kJ mol$^{-1}$of) the anisotropic model. There is no obvious reason for this jump.

| | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| Aniso | 2.92 | 2.80 | 2.74 | 2.72 | 2.71 |
| Iso | 2.90 | 2.86 | 2.75 | 2.71 | 2.68 |
| Iso nonpol+$\Delta Q$ | 2.86 | 2.85 | 2.83 | 2.82 | 2.84 |
| Iso nonpol+$\Delta Q/2$ | 2.93 | 2.91 | 2.89 | 2.89 | 2.89 |
| TIP4P/2005 | 2.77 | 2.78 | 2.74 | 2.75 | 2.74 |
| TIP4P/Ice | 2.78 | 2.80 | 2.76 | 2.75 | 2.76 |

Table 4.4: Separations $R_OO$ for the dimer and n-member rings given in Å. For a given model and ring, all bond lengths agreed to within 0.01 Å, so only average values are given.

| | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| MP2/CBS | −20.79 | −66.19 | −115.60 | −151.92 | −187.49 |
| Aniso | −21.013 | −61.310 | −116.866 | −154.049 | −189.939 |
| Iso | −20.567 | −55.804 | −107.423 | −146.881 | −184.190 |
| Iso nonpol+$\Delta Q$ | −29.259 | −77.327 | −129.540 | −170.189 | −192.448 |
| Iso nonpol +$\Delta Q/2$ | −22.653 | −60.028 | −100.317 | −131.159 | −160.898 |
| TIP4P/2005 | −28.712 | −76.984 | −128.277 | −167.324 | −204.246 |
| TIP4P/Ice | −31.745 | −85.057 | −141.808 | −185.083 | −225.974 |

Table 4.5: Binding energies in kJ mol$^{-1}$ for the dimer and n-member rings. MP2/CBS data from [133].

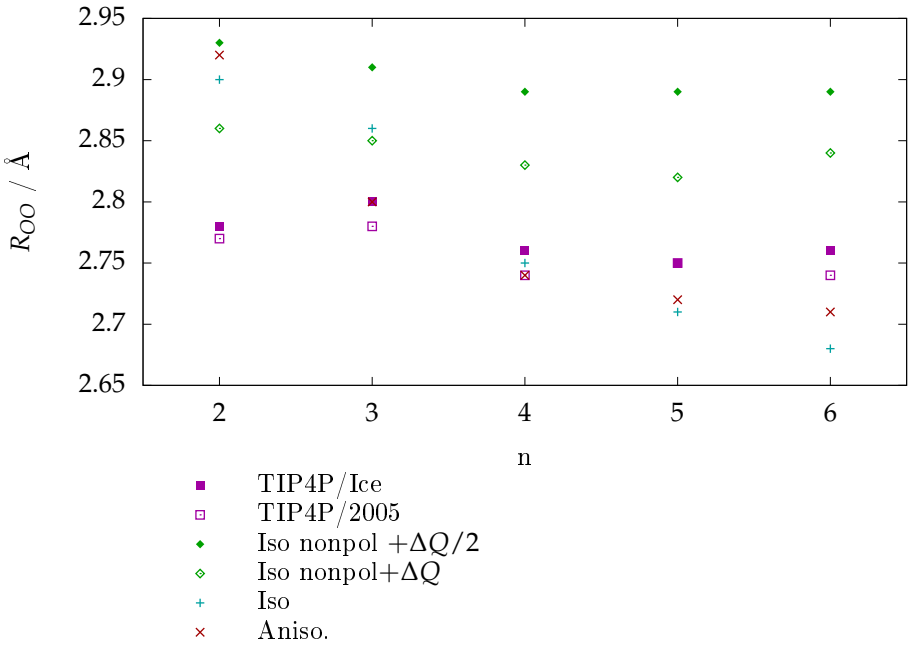
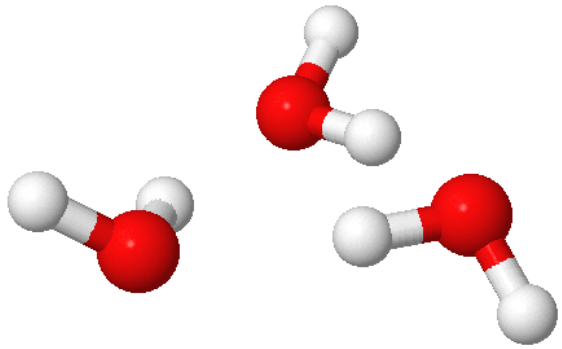
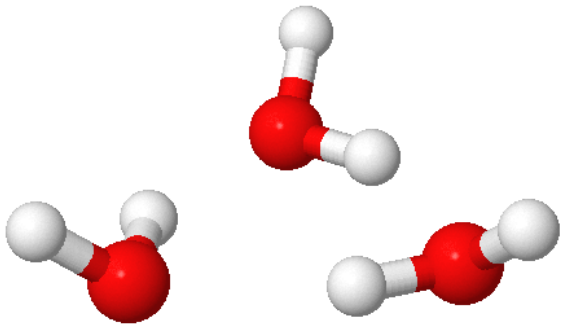



Figure 4.5: Average separations between neighbours $R_{OO}$ for n-member water rings.

.



(a)



(b)

Figure 4.6: (I) The correct, "up up down", minimum energy configuration for the water trimer. (II) The "up up up" configuration found to be the minimum using the anisotropic and isotropic polarisable potentials.

#### 4.2.1.2 Water hexamers

The hexamers looked are were, in order of increasing energy, the "prism", "cage", "book", "bag", and "ring" structures shown in Fig 4.8. The anisotropic model makes an error quite common among water models: The cage structure is incorrectly identified as the lowest energy hexamer, with the prism lying less than 0.16 kJ mol$^{-1}$ behind. The energy of the prism and relative energies of each hexamer are shown in Table 4.6, and energies for anisotropic models are plotted in Fig 4.9.

Of the models listed in Table 4.6 only the MB-Pol and SCME/GAP (this is the SCME model with the 3-body MB-Pol term included) models get the ordering of the hexamers correct, although in the case of SCME/GAP the difference is extremely small, less than 0.5 kJ mol$^{-1}$.

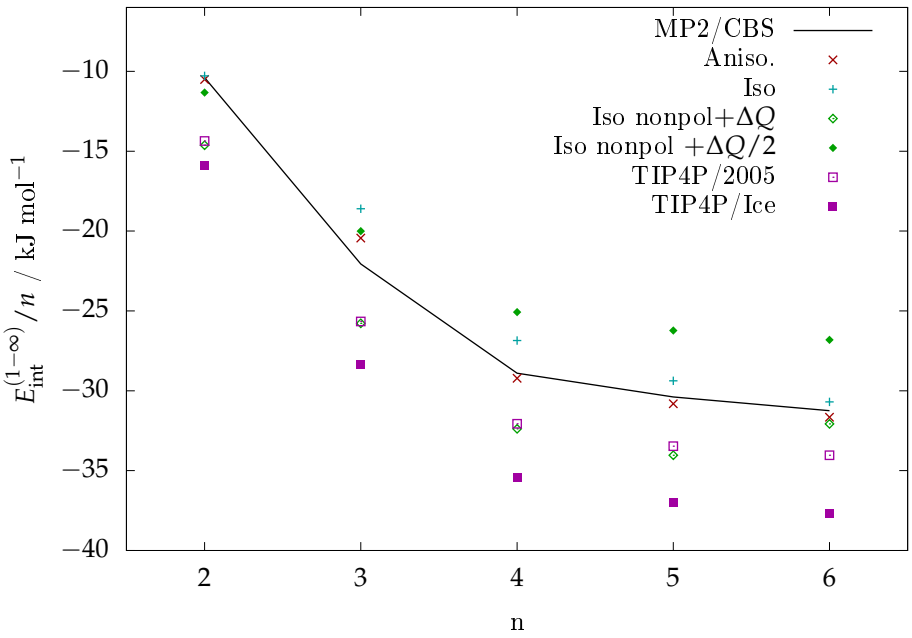
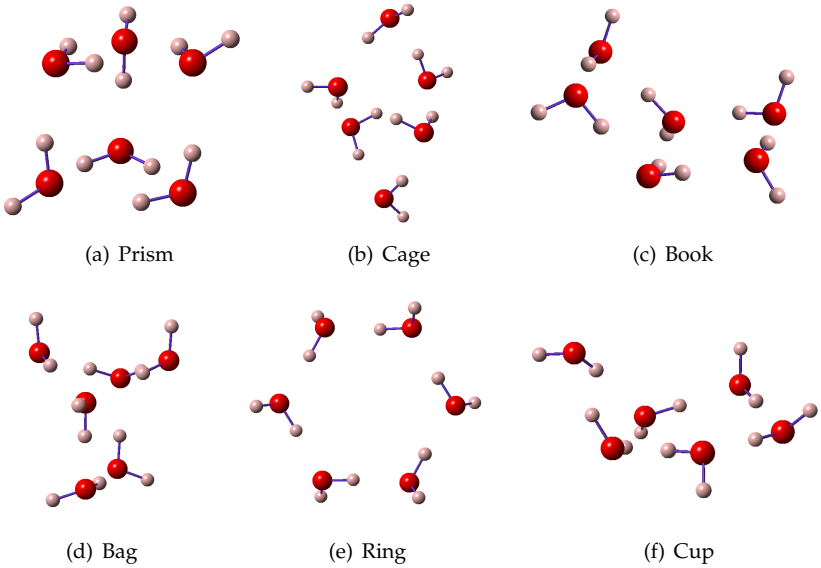



Figure 4.7: Energies per molecule for n-member water rings. MP2/CBS data from [133].

Comparing the relative energies across models, the anisotropic model does similarly well to SCME/GAP and might be seen as doing a better job than MB-Pol when the ring is considered. DDP2, which was fitted to the CCSD(T) reference energies with the same basis set, comes very close to matching those relative energies, whereas our model lies somewhere between the two sets of reference energies.

Of the isotropic models, it is surprising that the non-polarisable model has the correct ordering of hexamers but the polarisable model does not. The energy separations are far too large, like the TIP4P non-polarisable models, although those models both have incorrect ordering.

For the anisotropic model, a more detailed comparison was made with theory. The structures for the prism, cage and book(1) hexamers are shown in Fig 4.10 with O-O separations for our model and from MP2 and experimental results by Perez et al (2012) [135] shown in Table 4.7. It should be noted that the hydrogen arrangement of the cage Perez et al. use is slightly different from our arrangement, these are shown in Lusada and Leutwyler (2002) [136] as the du1 and uu1 arrangements respectively. Perez et al. found the uu1 arrangement to have an energy only 0.24 kJ mol$^{-1}$ higher than du1 and found experimental evidence for uu1. For all three structures, our model has average bond lengths which are shorter than both experimental and MP2 values, but only by less than 0.05 Å.



(a) Prism (b) Cage (c) Book

(d) Bag (e) Ring (f) Cup

Figure 4.8: Low-lying water hexamers.

| Structure | Prism | Cage | Book | Bag | Ring |
|---|---|---|---|---|---|
| CCSD(T)$^{(a)}$ | −189.34 | 0.59 | 3.26 | - | 6.36 |
| CCSD(T)$^{(b)}$ | −200.96 | 1.09 | 2.89 | 5.82 | 7.36 |
| MP2/CBS | −191.89 | 0.29 | 1.05 | - | 4.39 |
| Aniso | −195.569 | −0.157 | 1.613 | 3.965 | 5.630 |
| DPP2 | −189.20 | −0.13 | 3.14 | - | 7.32 |
| SCME/GAP | −198.991 | 0.335 | 0.879 | - | 7.113 |
| MB-Pol | −201.54 | 1.34 | 4.90 | 7.82 | 14.267 |
| TTM4-F | −183.22 | −8.16 | −9.92 | −6.95 | −5.44 |
| Iso | −184.031 | 2.178 | 1.232 | 1.013 | −0.159 |
| Iso nonpol+$\Delta Q$ | −211.160 | 3.137 | 9.389 | 10.038 | 18.712 |
| Iso nonpol +$\Delta Q/2$ | −175.768 | 2.467 | 8.771 | 3.263 | 14.970 |
| TIP4P/2005 | −216.517 | −1.490 | 4.477 | 1.762 | 12.271 |
| TIP4P/Ice | −239.390 | 21.383 | 27.350 | 24.635 | 35.144 |

Table 4.6: Binding Energies of the prism and relative binding energies of other water hexamers in in kJ mol$^{-1}$. DPP2 and ($a$) data from [118], SCME/GAP data from [137], MB-Pol, TTM4-F and ($b$) data from [138]. MP2/CBS data from [133]. ($a$) using aug-cc-pVTZ basis set optimised at MP2/aug-cc-pVTZ, ($b$) using VTZ-F12 basis set.

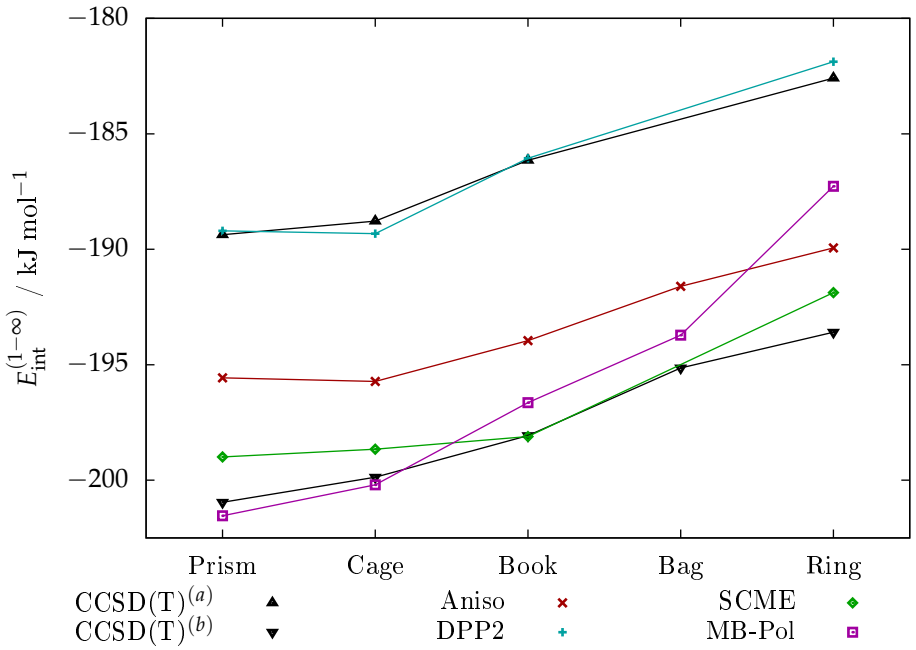



Figure 4.9: Hexamer energies for polarisable models and reference energies. DPP2 and (*a*) data from [118], SCME/GAP data from [137], MB-Pol, TTM4-F and (*b*) data from [138]. MP2/CBS data from [133]. (*a*) using aug-cc-pVTZ basis set optimised at MP2/aug-cc-pVTZ, (*b*) using VTZ-F12 basis set.

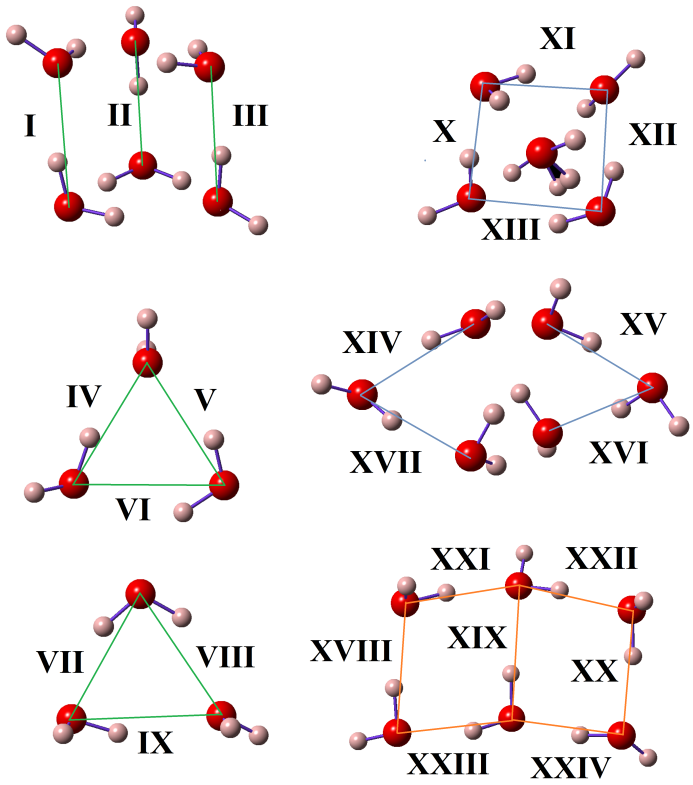



Figure 4.10: **I-IX** show distances measured for the prism hexamer, **X-XVII** show distances measured for the cage hexamer and **XIIX-XXIV** shows distances measured for the book(1) structure. Distances are given in Table 4.7

.

| Line | Aniso. model | Experiment | MP2 |
|---|---|---|---|
| I | 2.91 | 2.79 | 2.78 |
| II | 2.66 | 2.70 | 2.69 |
| III | 2.74 | 2.96 | 2.96 |
| IV | 2.77 | 2.93 | 2.92 |
| V | 2.87 | 2.81 | 2.81 |
| VI | 2.95 | 2.99 | 3.02 |
| VII | 2.83 | 2.98 | 3.00 |
| VIII | 2.90 | 2.83 | 2.83 |
| IX | 2.95 | 2.99 | 2.99 |
| Average (rmse) I-IX | 2.842 | 2.887 (0.044) | 2.889 (0.047) |
| X | 2.67 | 2.71 | 2.71 |
| XI | 2.89 | 2.96 | 2.96 |
| XII | 2.92 | 2.98 | 2.97 |
| XIII | 2.97 | 3.01 | 3.02 |
| XIV | 2.80 | 2.79 | 2.79 |
| XV | 2.72 | 2.79 | 2.78 |
| XVI | 2.75 | 2.75 | 2.76 |
| XVII | 2.78 | 2.82 | 2.82 |
| Average (rmse) X-XVII | 2.813 | 2.851 (0.048) | 2.851 (0.046) |
| XVIII | 2.78 | 2.98 | 2.98 |
| XIX | 2.92 | 2.81 | 2.81 |
| XX | 2.69 | 2.73 | 2.73 |
| XXI | 2.78 | 2.82 | 2.82 |
| XXII | 2.67 | 2.72 | 2.72 |
| XXIII | 2.81 | 2.83 | 2.84 |
| XXIV | 2.69 | 2.72 | 2.73 |
| Average (rmse) XVIII-XXIV | 2.762 | 2.801 (0.092) | 2.804 (0.093) |
| Average (rmse) I-XXIV | 2.809 | 2.850 (0.092) | 2.852 (0.093) |

Table 4.7: O-O separations in  for the prism, cage and book(1) hexamers as shown in Fig 4.10 for the anisotropic water model. **I-IX** show distances measured for the prism hexamer, **X-XVII** show distances measured for the cage hexamer and **XIIX-XXIV** shows distances measured for the book structure (cyclic book(2) for our model and book(1) for reference distances). Experiment and MP2 results from Perez et al. (2012) [135]. MP2 has vibrational perturbation theory applied with cc-pVTZ-F12 basis.

### 4.2.2 Methane Clusters

Two clusters of six methane molecules from basin hopping results with the anisotropic potential were reminimised using the isotropic and Lennard-Jones methane potentials, giving the clusters shown in Fig 4.11. There is no reference data so it is assumed for this section that the anisotropic potential gives correct results, in any case the subsequent models were intended to resemble it. Lengths and energies are given in Table 4.8, from which it is immediately apparent that there is no need for polarisation in the isotropic methane model. In the first cluster the isotropic and anisotropic models appear to match in the orientations of molecules in I and III, but in the Lennard-Jones model only line I is correct. For both clusters, the separations in the Lennard Jones model are a closer match to the anisotropic model than the isotropic model is. However, the Lennard-Jones model has poorer agreement of energies and of energy difference between the two structures.

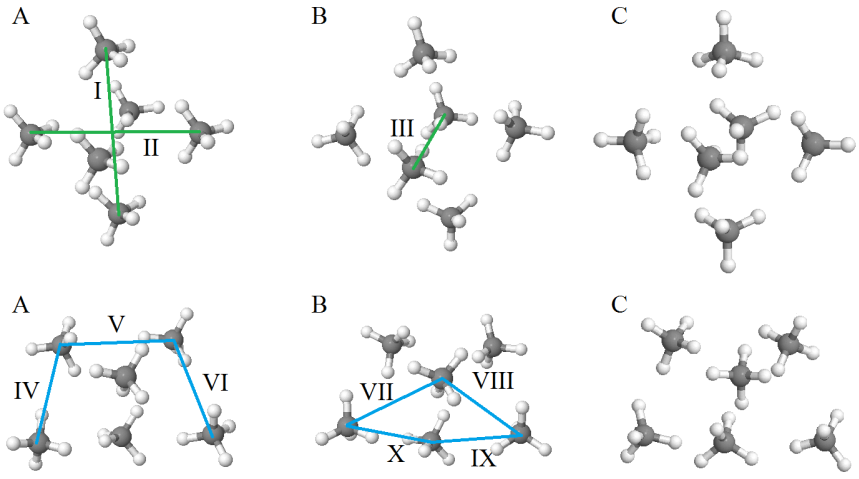



Figure 4.11: Two clusters of six methane molecules shown for three models: The anisotropic model (A), the isotropic nonpolarisable model (B) and the Lennard-Jones model (C). All distances measured agreed to within 0.01 when comparing the polarisable and nonpolarisable isotropic models, and there was no perceivable difference in the orientation of the molecules. The isotropic models appear to give very similar structures to the anisotropic model however the separations are better reproduced using the Lennard-Jones model. See Table 4.8 for lengths and energies.

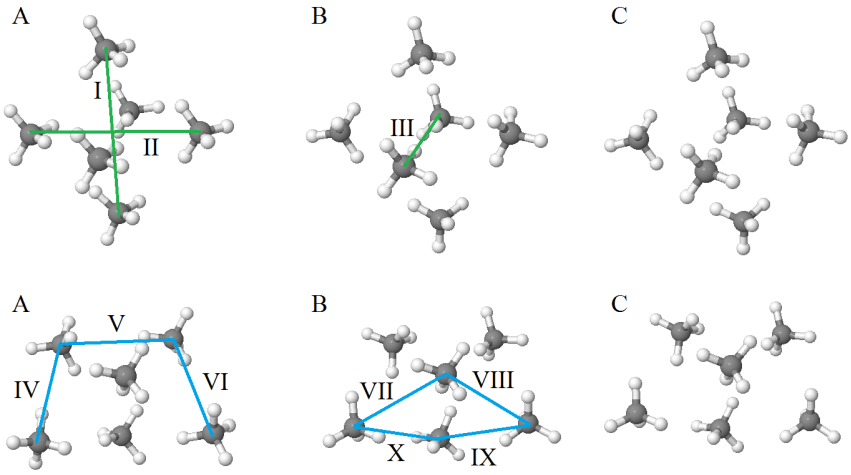



Figure 4.12: Two clusters of six methane molecules shown for three models: The anisotropic and polarisable model (A), the isotropic polarisable model (B) and the isotropic nonpolarisable model (C). All distances measured agreed to within 0.01 when comparing the two isotropic models, and there is no perceivable difference in the orientation of the molecules. See Table 4.8 for lengths and energies.

| | Anisotropic | Isotropic, polarisable | Isotropic, nonpolarisable | Lennard-Jones |
|---|---|---|---|---|
| Cluster 1 | | | | |
| Energy | -20.45 | −19.73 | −19.74 | −22.981 |
| I | 5.96 | 6.24 | 6.24 | 5.75 |
| II | 6.04 | 6.24 | 6.24 | 5.75 |
| III | 5.44 | 4.89 | 4.89 | 5.65 |
| Cluster 2 | | | | |
| Energy | -20.387 | −19.14 | −19.14 | −22.202 |
| IV | 4.17 | 4.14 | 4.14 | 4.02 |
| V | 4.15 | 3.96 | 3.96 | 4.05 |
| VI | 4.16 | 3.96 | 3.96 | 4.18 |
| VII | 3.95 | 4.13 | 4.13 | 4.01 |
| VIII | 3.96 | 3.92 | 3.92 | 4.01 |
| IX | 4.12 | 3.93 | 3.93 | 3.97 |
| X | 4.12 | 3.98 | 3.98 | 3.96 |

Table 4.8: Energies in kJ $mol^{-1}$ and distances in Å for the two clusters of methane shown in Fig 4.12, using different models. The non-polarisable model gives energies and measurements negligibly different from the polarisable model, indicating polarisability is not necessary. Using the anisotropic model as a reference, the Lennard-Jones potential performs better than the isotropic model with regards to C-C separations but worse with regards to energies.

### 4.2.3 Water-Methane Clusters

Methane clathrate is generally thought to require a guest molecule (methane) and some arrangement of (edge-sharing) pentamer or hexamer rings as the start point for nucleation. Walsh et al. [83] found that in particular the arrangement of two methane molecules projected in either direction perpendicular to the face of a water pentamer ring (Fig 4.13) was the starting point for nucleation, even in the formation of sI clathrate which does not contain such an arrangement. This is perhaps due to the relatively high number of pentamer rings compared to hexamer rings present in water.

With the anisotropic potentials, basin hopping searches using five water molecules and two methane molecules gave the the aforementioned ring arrangement as the lowest energy structure by 11.0 kJ mol$^{-1}$. This structure was reminimised for each of our water potentials, energies and average intermolecular separations are given in Table 4.9. In addition, the same process was carried out for a similar arrangement using two methane molecules and the hexamer ring (Fig 4.14), with the results given in Table 4.10. The methane potentials used were the anisotropic and the non-polarisable isotropic potentials.

In both cases the polarisable models give an oxygen-oxygen separation close to the value found without including the methane molecules, whereas the non-polarsiable model has increased the separation by 0.07 and 0.15 Å for the water pentamer and hexamer rings respectively. The $R_{CC}$ distance, which is not close to the methane dimer, is lower by about 0.8 Å for both structures using the isotropic model. The separation for the non-polarisable model was 1.1 Å lower than the anisotropic case for five water molecules and 3.2 Å greater for six molecules. This difference might be attributed to the shallowness of the methane-methane and methane-water potentials. The non-polarisable model also has these two structures energetically in the wrong order.

| | $R_{OO}$ Å$^{-1}$ | $R_{CC}$ Å$^{-1}$ | Energy / kJ mol$^{-1}$ |
|---|---|---|---|
| Aniso | 2.72 | 6.04 | $-169.640$ |
| Iso | 2.71 | 5.19 | $-170.159$ |
| nonpol $+\Delta Q$ | 2.71 | 5.33 | $-207.150$ |
| nonpol $+\Delta Q/2$ | 2.89 | 4.91 | $-154.769$ |

Table 4.9: Average separations and energies for the cluster of five water molecules and two methane molecules shown in Fig 4.13.

| | $R_{OO}$ Å$^{-1}$ | $R_{CC}$ Å$^{-1}$ | Energy / kJ mol$^{-1}$ |
|---|---|---|---|
| Aniso | 2.71 | 5.33 | $-207.510$ |
| Iso | 2.68 | 4.55 | $-211.418$ |
| nonpol $+\Delta Q$ | 2.82 | 5.46 | $-188.960$ |
| nonpol $+\Delta Q/2$ | 2.98 | 8.53 | $-150.713$ |

Table 4.10: Separations and energies for the cluster of six water molecules and two methane molecules shown in Fig 4.14.

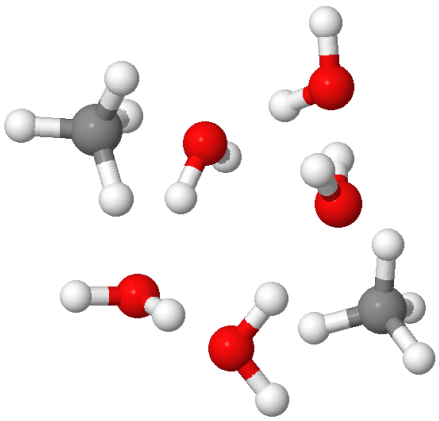


Figure 4.13: Minimum energy cluster of five water molecules and two methane molecules

.

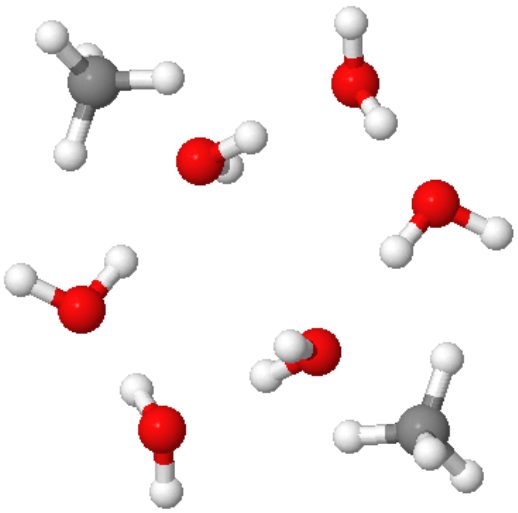


Figure 4.14: Cluster of six water molecules and two methane molecules

.

## 4.3 Second Virial Coefficients

The second virial coefficient $B(T)$ is given by

$$B(T) = -\frac{1}{2}\int\int(e^{E_{\text{int}}^{(1-\infty)}/kT} - 1)d\Omega dr^3 + \frac{\hbar^2}{24(kT)^3}(\frac{\langle \mathbf{F}^2\rangle_0}{M} + \sum_\alpha \frac{\langle \mathbf{T_{ff}}^2\rangle_0}{I_{\alpha\alpha}}) \tag{4.1}$$

where the first term above is the classical result $B(T)_{\text{Cl}}$ from integrating the Mayer function (the integration here is over separations and orientations) and the second term gives the quantum correction. Here $\langle \mathbf{F}^2\rangle_0$ and $\langle \mathbf{T_{ff}}^2\rangle_0$ are the mean square force and components of mean square torque on the molecule respectively and $I_{\alpha\alpha}$ are the molecule's moments of inertia.

The separations used for the integration are at specified intervals (chosen as 0.2 Bohr) between two points. The models which used damped polarisation or no polarisation at all used separations between 0.2 and 20 Bohr; for the models which used undamped polarisation it was necessary to limit the separations to between 4 and 20 Bohr.

The orientations are described by Euler angles, there are six Euler angles for two molecules but one angle is kept at zero. The remaining angles are varied according to a Sobol sequence so as to give a uniform distribution of orientations. For a given temperature, the integration of the Mayer function over $\Omega$ is carried out first at each separation, then the integration over separations takes place.

The number of orientations used is specified to a power of 2. Orient starts with 1024 orientations and increases the number of orientations by factor of 2, recalculating $B(T)$ at each step. This allows for specification of a large number of orientations with the option of stopping the procedure early if $B(T)$ has converged. The number of orientations used varied according to the model, with the lowest number being 32768.

### 4.3.1 Second virial coefficient for water

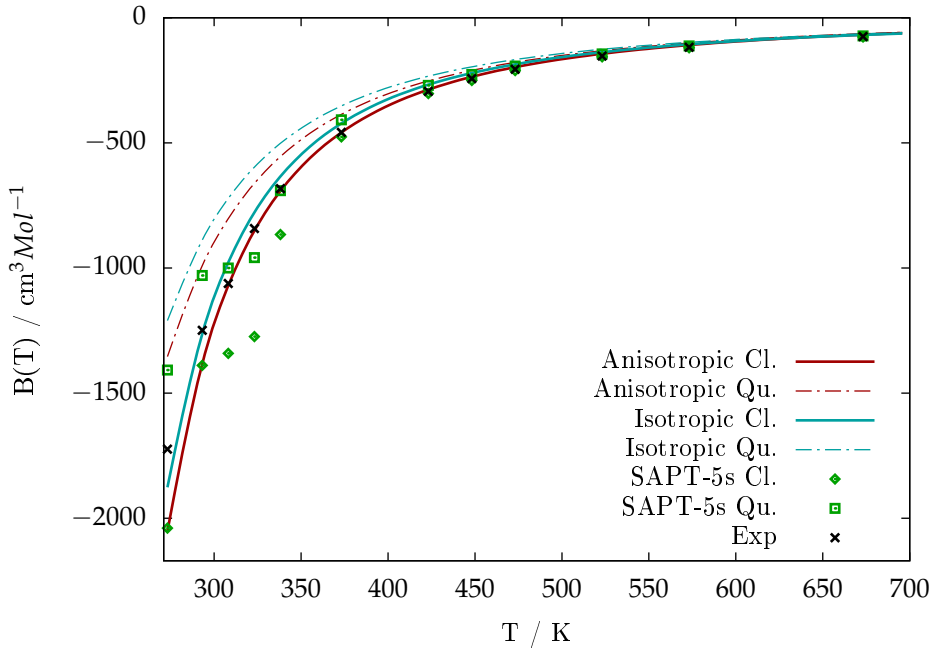
The second virial coefficient for water here is plotted alongside SAPT-5s and experimental data taken from Mas et al. (2000) [14]. At the classical stage the anisotropic potential matches the experimental data very closely, however when the quantum correction is included the SAPT-5s potential gives results closer to the experimental data. The charge model which uses the $+\Delta Q/2$ adjustment performs similarly well to the polarisable isotropic model, whereas the charge model which uses the $+\Delta Q$ adjustment performs poorly.

### 4.3.2 Second virial coefficient for methane

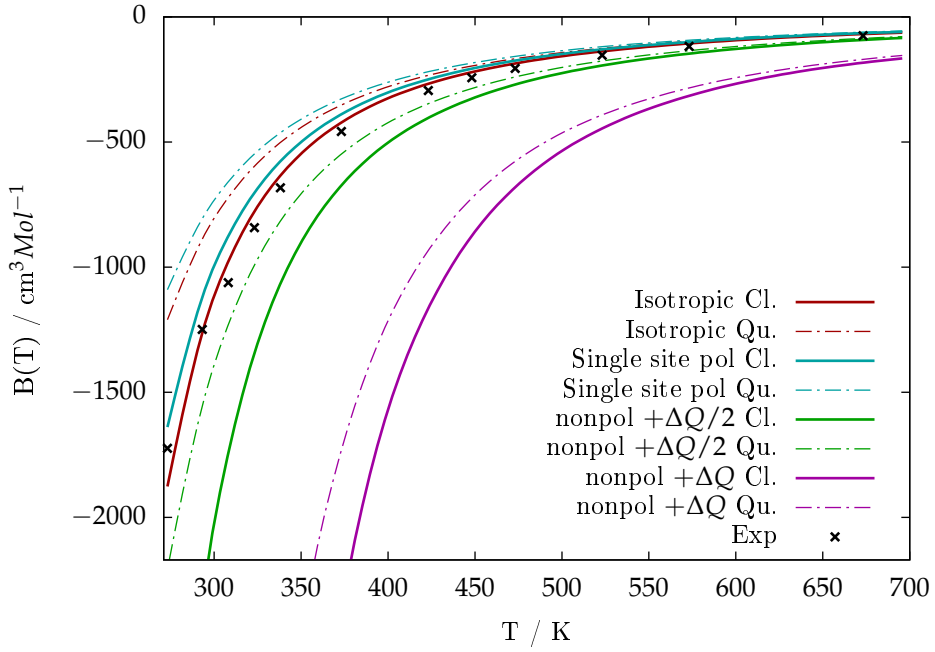
The virial coefficient as found using the anisotropic, isotropic non-polarisable, and Lennard Jones methane models. All models perform similarly well when compared with experimental data.

### 4.3.3 Second virial coefficient for water-methane

The second virial coefficient for water-methane is compared with both experimental and CCSD(T) results taken from Akin-Ojo and Szalewicz (2005) [40]. The anisotropic potential matches the CCSD(T) results closely, although all potentials give values higher than experiment.




(a)




(b)

Figure 4.15: Second virial coefficient for water using the anisotropic and isotropic models (4.15(a)) and simplifications of the isotropic model (4.15(b)) plotted against temperature. SAPT-5s and experimental data taken from Mas et al. (2000) [14].

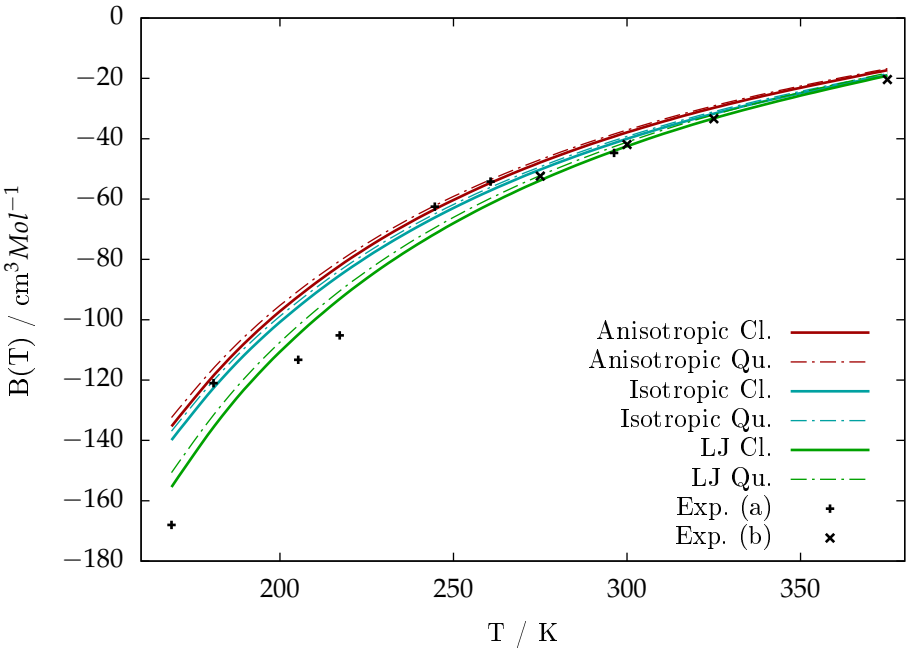



Figure 4.16: Second virial coefficient for methane plotted against temperature for models at different stages of simplification, alongside experimental data taken from (a) [139], (b) [140].

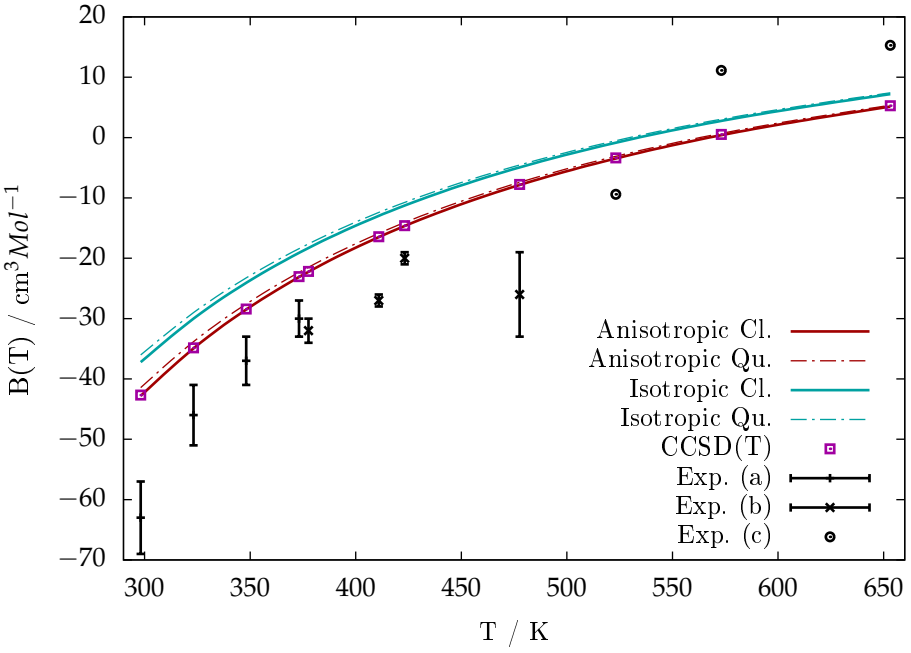



Figure 4.17: Cross second virial coefficient for water-methane plotted against temperature for models at different stages of simplification, alongside CCSD(T) and experimental data taken from [40].

## 4.4 Estimating the melting temperature of water models

One notorious difficulty with water models is the tendency to underestimate the melting point of ice Ih. As an example, the POL4D model which contains 5 charge sites and a polarisable oxygen site, and was parameterised with the intent of establishing the correct melting temperature, was only able to achieve a melting temperature of 260 K [141]. Of the commonly used models only TIP4P/Ice, parameterised to reproduce the phase diagram for solid phases of water, gives a melting temperature for ice Ih of 270 K. The TIP5P model is also able to give a very good melting point of ice Ih at 272 K, however it is known that using TIP5P ice II is more stable than ice Ih at this temperature [129].

Estimates of melting temperature were made by simulating an ice-water interface and judging by eye if any melting had occurred. This was not the intended method, see Section 5.4 for details. To prepare the cell, the start point was a cell of ice Ih $8\times12\times8$ unit cells in size (6144 molecules), giving dimensions of approximately $50\times50\times60$ Å throughout the simulation. This cell was divided into two along the z-axis; then half the cell was kept 'frozen' in place while the other was melted by simulating at constant volume with a starting pressure of 1 atm and at a temperature of 310K for 100 ps; this used the $+\Delta Q/2$ model. Simulations used a Nose-Hoover thermostat with a relaxation time of 0.4 ps; and a step size of 0.2 fs. The choice of step size is consistent with previous clathrate simulations, for example Bagherzadeh et al. [72, 96] who use the TIP4P/Ice water model. The choice of relaxation time must be large enough to allow for natural fluctuations to take place but must ensure that these fluctuation times are small on the scale of the simulation time. The relaxation time was intended to be used for all simulations; which run from the order of 100 ps to 10 ns. Comparing these to the 0.2 fs step size, 0.4 ps can be justified in so far as it is of an appropriate order of magnitude.

With the cell now containing one half ice and one half liquid water, the cell was allowed to melt under NPT conditions at 1 atm and a chosen temperature. The barostat relaxation time has similar requirements to the thermostat relaxation time; i.e it must allow for small fluctuations but not drift. Although temperature and pressure fluctuations do not necessarily occur at the same rate, the previous considerations about simulation time and step size still apply and in the interest of simplicity the barostat relaxation time was set to 0.4 ps, matching the thermostat. The simulations ran until melting was very obvious, as shown in 4.18, with a maximum time of 2 ns. Estimates using this rough method are given in Table 4.11, alongside melting temperatures for other models using the previously described method. None of the charge models perform as well as the TIP4P/Ice or TIP4P/2005 models, but the $+\Delta Q$ model at a melting temperature between 230 K and 240 K is much closer to the actual value than the $+\Delta Q/2$ model, which melts between 140 K and 150 K, worse than any of the existing models looked at. The polarisable shell model had a lower melting temperature still, below 120 K.

## 4.5 Diffusion Coefficient

### 4.5.1 Self-diffusion coefficient of water

For simulations of water, cell used contains 6144 molecules and was made by melting a cell of ice Ih $8\times12\times8$ unit cells in size, giving dimensions of approximately $50\times50\times60$ Å throughout the simulation. This is a relatively large cell, which should reduce the diffusion

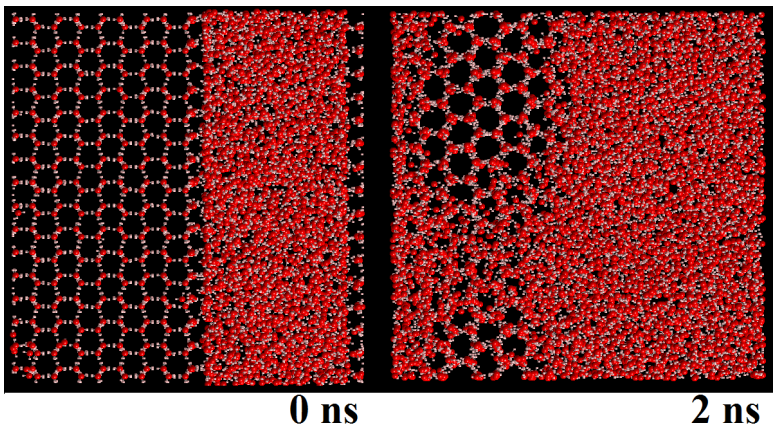



Figure 4.18: 5.2(a) Cell at the start of a simulation for determining melting point, and after 2 nanoseconds at 170 K and 1 atm for the $+\Delta Q/2$ model, showing that the ice has melted.

| Model | Melting temperature |
|---|---|
| nonpol $+\Delta Q$ | 230-240 |
| nonpol $+\Delta Q/2$ | 140-150 |
| polarisable oxygen | 100-120 |
| SPC/E[142] | 213 |
| TIP4P[142] | 229 |
| TIP4P/2005[142] | 250.5 |
| TIP4P/Ice[142] | 270 |
| TIP5P[142] | 272 |
| 6-site NdvE[143] | 290 |
| POL4D[141] | 260 |

Table 4.11: Melting temperature in K for ice Ih, from simulation of an ice Ih/liquid water interface.

coefficient's dependence on cell size, observed by [144] for the TIP3P model. All simulations used NPT conditions through a Nose-Hoover thermostat and barostat each with a relaxation time of 0.4ps, and a timestep of 2 fs. The diffusion coefficient was found by plotting the mean-square displacement against time; the plot quickly approaches a straight line and the diffusion coefficient is given by one sixth of the gradient.

The coefficients at ambient conditions are given in Table 4.12 alongside other models and experimental results. The diffusion coefficient of the $+\Delta Q$ model has a similar value to both the TIP4P/2005 model and the experimental result of 0.23 $\text{Å}^2$ $\text{ps}^{-1}$. The diffusion coefficient for the $\Delta Q/2$ model is more than three times this value at 0.78 $\text{Å}^2$ $\text{ps}^{-1}$, and the polarisable oxygen model lies between these two but closer to the latter at 0.59 $\text{Å}^2$ $\text{ps}^{-1}$.

| | Experiment | MB-Pol | TIP4P/2005 | $+\Delta Q$ | $+\Delta Q/2$ | Opol |
|---|---|---|---|---|---|---|
| D | 0.23 | 0.12 (cl), 0.22 (qu) | 0.249 | 0.25 | 0.78 | 0.59 |

Table 4.12: Diffusion coefficient in $\text{Å}^2$ $\text{ps}^{-1}$ at 1 atm and 298.15 K, except for TIP4P/2005 data from [145] at 300 K and the infinite cell size limit. Experimental and MB-Pol data from [146].

### 4.5.2 Self-diffusion coefficient of methane

For the methane models, the previous method was used with 8000 molecules, which took up a cell around 100 Å wide. The temperatures and pressures chosen were those from reference data that most closely matched the conditions of later clathrate simulations. The UAM model gives results accurate to within around 5% at both temperatures. The isotropic model developed here has a diffusion coefficient too low by 15% at 303.3 K and too high by around 7% at 333.1 K. The Lennard-Jones type model fares slightly better, getting the diffusion coefficient too low by around 12% at 303.3 K and by only 3% at 333.1 K.

| | Experiment | UAM | Lennard-Jones | Iso |
|---|---|---|---|---|
| 303.3 K | 6.2 | 6.12 | 5.46 | 5.27 |
| 333.1 K | 7.5 | 7.87 | 7.27 | 8.05 |

Table 4.13: Diffusion coefficient in $Å^2$ $ps^{-1}$ at 30 MPa (296.1 atm) and a given temperature for different methane models. Experimental data from [147].

## 4.6 Radial Distribution Functions

The previous simulation setup was also used to find radial distribution function (RDF) data for liquid water and for ice Ih. For liquid water, comparisons can be made with experimental data by Soper and Benmore [148], and by Skinner et al. [149]. For ice Ih, comparison is limited to other water models, in this case TIP4P/Ice using the same simulation conditions.

### 4.6.1 Liquid Water

The O-O, O-H and H-H radial distribution functions for liquid water are given in Fig 4.19.

The O-O distribution for the $+\Delta Q$ model matches the experimental data well, whereas the $+\Delta Q/2$ and polarisable oxygen models are almost featureless after the first peak. The first peak appears around 0.05 Å too far out for $+\Delta Q$ and around 0.15 Å too far out the $+\Delta Q/2$ and polarisable oxygen models when compared with experiment. The $+\Delta Q$ model does place the first peak too high by about 0.3, which the other models agree on the peak's height to within 0.1.

For the O-H distribution, all models clearly have too steep an approach to the first peak. The $+\Delta Q$ model places the first two peaks around 0.05 Å too far out and too high by around 0.30 and 0.15 for the first and second peaks respectively. The first trough is 0.10 Å too far out and about 0.02 too high. The third peak is also too far out, by about 0.2 Å. The $+\Delta Q/2$ model has the first two peaks out by around 0.15 Å and the third peak out by around 0.5 Å. The three peak heights all agree with the experimental data to within 0.05. The first trough is around 0.16 Å too far out and around 0.1 too high. The polarisble oxygen model gives a similar RDF to the $+\Delta Q/2$ model but its peaks and troughs are slightly higher, by around 0.03, and the $R_{OH}$ distances for the first trough and second peak are slightly lower, by around 0.06 Å.

The H-H distribution again shows all models rising too steeply to the first peak. The first peak is again slightly too far out, around 0.05 Å for the $+\Delta Q$ model and 0.18 Å for the $+\Delta Q/2$ and polarisable oxygen models. The $+\Delta Q$ model also places the peak too high by around 0.15 and the polarisable model by around 0.05. The first trough is also too far out by around 0.12 Å

for the $\Delta Q$ model and around 0.27 Å for the other models. All models place the trough too high, by around 0.15 for $+\Delta Q$, 0.30 for $+\Delta Q/2$ and 0.33 using polarisable oxygen. The second peak isn't very sharp and so its position is harder to determine, but the $+\Delta Q$ and polarisable models seem to get its position about right whereas the $+\Delta Q/2$ model has it around 0.07 Å too far out. The peak for $+\Delta Q$ is at about the right height, but is too low by around 0.04 for the $+\Delta Q/2$ model and 0.33 for the polarisable oxygen model.

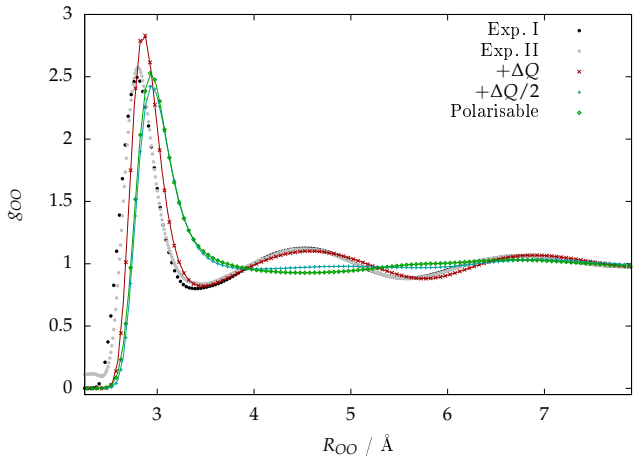
### 4.6.2 Ice Ih

Reference data for ice Ih is not readily available, so instead the models are compared with the TIP4P/Ice model, which is seen as a good model for ice Ih and generally gets the same features for the OO distribution found in experiment.

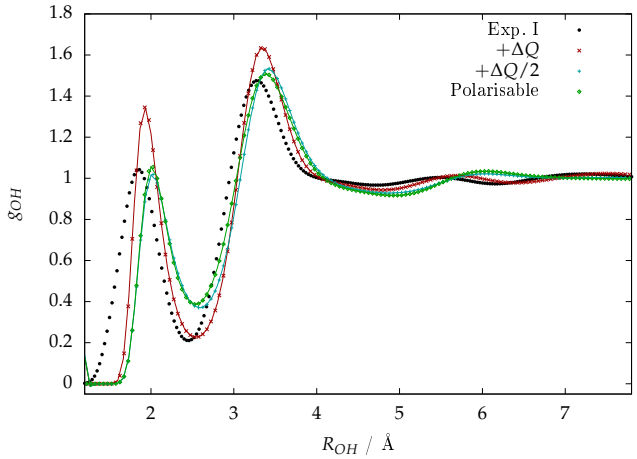
At first, the simulations to find the radial distribution functions were run at 230 K. At this temperature the oxygen-oxygen distribution for $+\Delta Q/2$ model does not resemble ice Ih, for this reason the simulation was repeated at 100 K, see Fig 4.20(a). The polarisable model had sharper peaks at 100 K compared with 230 K but was not qualitatively different.

The OO distribution for the $+\Delta Q$ model is very similar to the result using the TIP4P/Ice model but stretched out slightly; the first peak is 0.06 Å out and the second peak is around 0.15 Å out. At 230 K the $+\Delta Q/2$ model has only one peak at about 2.9 Å, but at 100 K it has the same general shape with first and second peaks 0.15 Å and 0.20 Å further out that the TIP4P/Ice model. The polarisable model has it's first peak further in, about 0.2 Å less than TIP4P/Ice, with the second peak appearing 0.3 Å before TIP4P/Ice, and the shoulder which should appear about 0.9 Å later is absent.

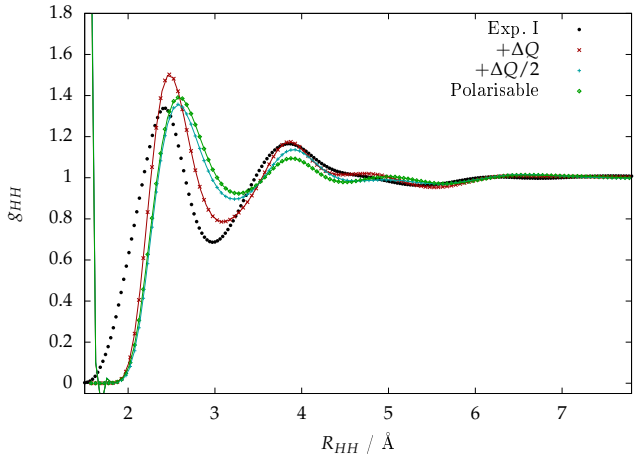
The OH and HH distributions follow this same general trend: the $+\Delta Q$ and $+\Delta Q/2$ (at 100 K) models get the general shape right but have the peaks slightly further out, with $+\Delta Q/2$ further out than $+\Delta Q$. The polarisable model has peaks closer together and is generally lacking in features.




(a)




(b)




(c)

Figure 4.19: Radial distribution functions for water at 298.15 K and 1 atm using different charge models, compared with experimental results at 300 K and 1 atm. The sharp peak seen at around 1.5 Å is from hydrogens in the same molecule. Exp. I from [148], Exp. II from [149].

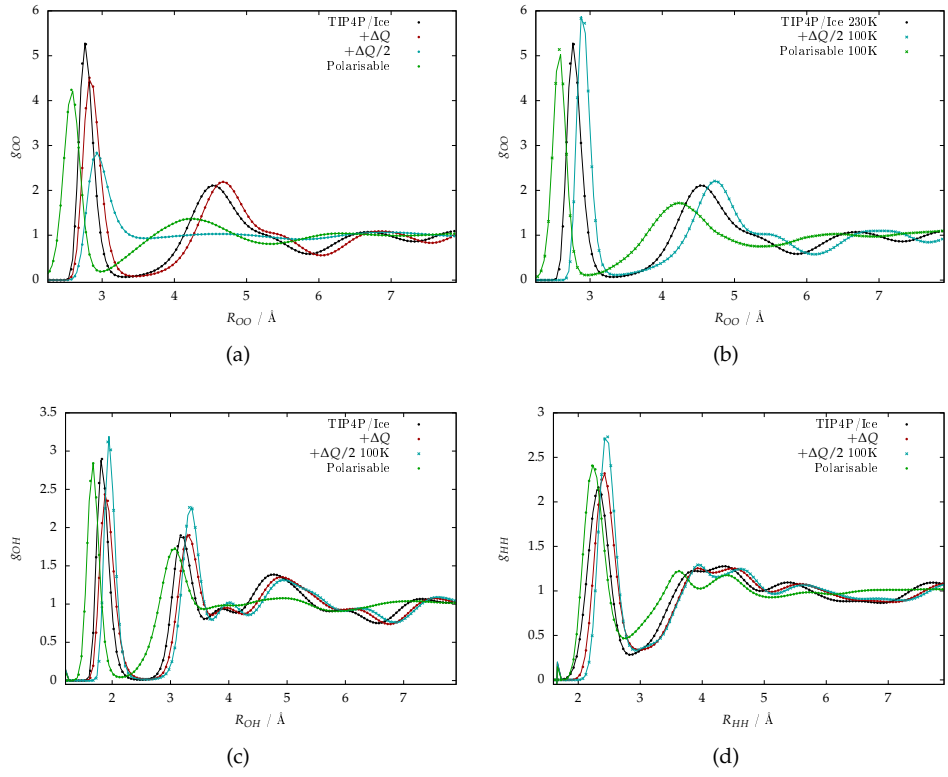



Figure 4.20: Radial distribution functions for OO in ice Ih at 1 atm and 4.20(a) 230 K, 4.20(b) 100 K; and OH 4.20(c) and HH 4.20(d) RDFs at 1 atm and 230 K.

## 4.7 Discussion

### 4.7.1 The anisotropic water model

The anisotropic water model is able to reproduce the minimum-energy dimer very well, better than other models for which minimum energy geometries were readily available including the polarisable CC-Pol model. This model also predicted the energetic ordering of low-lying hexamers and relative energies similarly well to comparable models except for the Book structure in which our model has a relative energy over 1 kJ mol$^{-1}$ too low. Our model is also slightly closer in total energy for the prism structure than the SCME or MB-Pol models, but not the DPP2 or TTMF-4 models. It should be noted the TTMF-4 model has an energetic ordering of no clear resemblance to reference energies. The relatively poor performance of the book structure may relate to the failure of the model to reproduce the non-bonding hydrogen positions in the structure book(1).

Overall the structures of the three lowest-energy hexamers are in close agreement with ab initio and experimental structures, with oxygen-oxygen separations generally being within 0.05 Å of reference lengths but lower in 20 of the 24 separations. The separation for the dimer is actually slightly higher (0.08 Å) than for the highest-level reference energies, which might suggest the closer separation is a result of the polarisation model, however the difference is small and this is far from certain.

The second virial coefficient using this model was in excellent agreement with experimental data but was consistently too high (in the sense of being too positive) when the quantum correction was applied.

### 4.7.2 Simplifying the water model

At the first stage of simplification, to a charge-only model with polarisable dipoles on each atom site, the minimum-energy dimer kept close to the correct separation but the H-O-O angle $\theta$ became too small and the O-O-bisection angle $\phi$ too large, similar to the results found with TIP4P models. The depth of the minimum also fell by around 0.4 kJ mol$^{-1}$ to -20.6 kJ mol$^{-1}$.

Replacing the polarisation with altered charge models further reduced $\alpha$ to less than 0.5° and gave values of $\phi$ which were less accurate than either TIP4P models. Compared with the isotropic model, the $+\Delta Q$ charge model reduced $R_{OO}$ by about 0.03 Å and increased $\phi$ by about 12°; whereas the $+\Delta Q/2$ model increased $R_{OO}$ by 0.03 Å and decreased $\phi$ by around 15°. Most startling is the effect of the charge model on the energy of the minimum: The $+\Delta Q$ model had a depth of -29.3 kJ mol$^{-1}$, this is close to the depths of the TIP4P models and far from the actual depth or the depth of the polarisable models. The $+\Delta Q/2$ model had a depth of -22.6 kJ mol$^{-1}$, only 2.1 kJ mol$^{-1}$ lower than the polarisable model.

As with the dimer energy, the $+\Delta Q/2$ model gives a second virial coefficient much closer to the experimental result and to the isotropic model.

The hexamer ordering for the isotropic model is entirely incorrect and the ring hexamer is predicted as the lowest energy structure. Despite this, the $+\Delta Q$ model yields the correct energetic ordering and the $+\Delta Q/2$ model has the correct ordering except for the bag, which is given an energy between the cage and the book. This suggests that the type of polarisation employed, i.e. a scaled undamped dipole on each atom site, is insufficient to describe polari-

sation of water clusters. In general the non-polarisable models have larger relative energies, with the $+\Delta Q$ model lying between the TIP4P/Ice and TIP4P/2005 energies, and the $+\Delta Q/2$ model having the lowest relative energies. The energies of water clusters and water-methane clusters tend to be too low for the $+\Delta Q$ model and too high for the $+\Delta Q/2$ model.

The radial distribution functions for ice and liquid water, and the estimates of the melting temperature, both indicate that the $+\Delta Q$ model is able to represent ice in simulation and the $+\Delta Q/2$ model is not. In general, the model with a single polarisable site gives similar radial distribution functions to the $+\Delta Q/2$ model, with an even lower melting point. It is known that the polarisation energy for this model is too low. From these points, and the similarities of the $+\Delta Q$ model to the TIP4P models which are well-established as successful models of bulk water, it can be inferred that the $+\Delta Q$ method is better than the $+\Delta Q/2$ method for the purpose of MD simulations.

### 4.7.3 The anisotropic methane model

There are no methane models in the literature which are of comparable complexity to the anisotropic methane model, making it difficult to evaluate. In comparison with all other models, the anisotropic model is distinctive in its superior match to the global minimum, although it placed the minimum around 0.2 Bohr out from the SAPT(DFT) reference energies.

### 4.7.4 Simplifying the methane model

The methane model was simplified to a non-polarisable point charge model using an exponential exchange-repulsion wall and damped $C_6$ dispersion (polarisation was found to be unnecessary, even in clusters of six molecules); and then to a model with only charges and Lennard-Jones terms. Using the isotropic model, the plot along the profile of both the global minimum energy dimer was shallower than the anisotropic model by 0.06 kJ mol$^{-1}$. Using the Lennard-Jones model, the depth was around 0.34 kJ mol$^{-1}$ shallower than the anisotropic model. For the secondary minimum, the Lennard-Jones model was close to the correct depth but reached the repulsive wall too early, whereas the isotropic model had the right general shape but was too shallow.The plot of the second virial coefficient gives an isotropic model matching the anisotropic model extremely closely.

For both clusters looked at, the isotropic model had too high an energy and the Lennard-Jones model too low an energy when compared with the anisotropic model. There were no obvious trends with regards to the bond lengths or orientations of the molecules, and there is no clear reason from the clusters to favour one isotropic model above the other.

Neither of the simplified models were able to match the diffusion coefficient as well as the single-site UAM model. Although the UAM model successfully predicts this bulk property, it has a potential very much more positive than either the first or second minimum profiles. This suggests an initial assumption made; that the methane model should be able to find something at least close to the global minimum when in the bulk; is incorrect. Any similar methane potentials produced in the future; i.e. any isotropic charge-only methane models; may fare better if the minimum is given no consideration at all during the parameterisation process.

# Chapter 5

# Simulations

This chapter describes MD simulation of clathrates; which was an end goal to the potential development. Simulations are carried out using both the models developed in Chapter 3 and the commonly-used combination of TIP4P/Ice and the United Atom Methane model. They serve as a test of the new models but also aim to give some insights into the clathrate decomposition process.

This chapter also gives details on the simulations described in Chapter 4, in particular those used to find the diffusion coefficient and the melting temperature for different charge models, along with a summary of the difficulties which led to developing a non-polarisable water model at all.

## 5.1 Details applied to all simulations

All simulations used the Velocity Verlet scheme with a Nose-Hoover approach used for both the thermostat and barostat.

The Velocity Verlet scheme calculates the velocities in two stages: Velocities are found after half a timestep, given by

$$\mathbf{v}(t + \Delta t/2) = \mathbf{v}(t) + \frac{\mathbf{F}(t)}{m} \tag{5.1}$$

for timestep $\Delta t$ and force $\mathbf{F}$. These are used to find the positions $\mathbf{r}$ at time $t + \Delta t$.

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \Delta t \mathbf{v}(t + \Delta t/2) \tag{5.2}$$

$\mathbf{v}(t + \Delta t)$ is then found based on $\mathbf{v}(t + \Delta t/2)$ and the forces calculated at from positions $\mathbf{r}(t + \Delta t)$:

$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t + \Delta t/2) + \frac{\mathbf{F}(t + \Delta t)}{m} \tag{5.3}$$

In the Nose-Hoover thermostat each particle interacts with a "virtual particle" which acts as a heat bath. This introduces an additional term to Newton's second law:

$$\frac{d\mathbf{v}(t)}{dt} = \frac{\mathbf{F}}{m} - \chi(t)v(t) \tag{5.4}$$

where thermostat friction coefficient $\chi$ depends on the target energy (i.e. target temperature) and on a specified "relaxation time" $\tau_{\mathrm{T}}$.

The barostat takes a similar approach using a barostat friction coefficient $\eta$ for constant pressure simulations, or a matrix of coefficients for simulations under constant stress. Again these coefficients change at a rate dependent on the target pressure (or stress tensor) and some relaxation time $\tau_P$. There are also interaction terms between the temperature and pressure coefficients.

The thermostat and barostat each change the velocity before and after the Velocity Verlet steps. As the kinetic energy and pressure depend on the velocity of the particles in the system, these must be evaluated before and after they change the velocity. The scheme used is to adjust the thermostat, then barostat, then thermostat again before and after each Velocity Verlet step of $\Delta t$, as set out by Martyna et al. [150]. For this reason $\eta$ must be evaluated at every $\Delta t/4$, and $\chi$ must be evaluated every $\Delta t/8$. For simulations at constant temperature but not constant pressure, $\chi$ is evaluated every $\Delta t/4$, this is described in the same paper.

For all simulations, $\tau_T$ and (if applicable) $\tau_P$ were set to 0.4 ps. For all simulations not containing a shell model, $\Delta t$ was set to 0.002 ps, this was reduced for the shell model and details of this are given where relevant.

## 5.2 Difficulties with Polarisable Potentials in DL_POLY

With the increased interest in polarisation in water, one of the initial aims was to study differences in behaviour between clathrates using polarisable and non-polarisable water potentials. The initial water potential is of a form more elaborate than those typically used in simulations; although they may become usable in the near future in OpenMM[151], for example.

To make a potential which could be used in DL_POLY, the most significant compromises made related to the polarisation model. As previously discussed, the best replacement for the electrostatic multipole model had seven charge sites in total, and this had to be abandoned because of the lack of available damping for the polarisation model.

Although a model was developed using undamped induced point dipoles at each atom site, this could not be transferred successfully to DL_POLY4. It is possible that this is because of the much closer approach between shells on hydrogen and oxygen atoms.

Subsequently the model was reduced to having only a single polarisable site at the location of the oxygen atom. At first it was assumed that the existing model could work well enough with only the polarisation model changed, or with dispersion terms altered slightly to readjust to the correct minimum depth, but this was fruitless and it was concluded that the models with three polarisable sites and with one polarisable site were too different; instead the repulsive wall parameters were refitted.

Further difficulties related to the nature of shell models: Firstly, there can always be some difference in energy expected when a charged site with an induced point dipole is replaced with two separate charges, this difference should be negligible if the core-shell separation is sufficiently small. Secondly, it is common in shell models that the shell represents the electrons and so all other intermolecular interactions occur between shells (or shells and charges). For our own purposes this is not desirable; it would be best to make interactions happen between cores since they are at the position for which the potential has been fitted. But again, this effect should be negligible if the core-shell separation is small enough.

In order to minimise the core-shell separation, naively the relationships to consider are that for a predetermined dipole moment the separation is inversely proportional to the (shell) charge, and this fixes the spring constant for a given polarisability.

To test this simulations were run in DL_POLY using only a single dimer which was minimised in a sufficiently large box, a cube 100 Å in each dimension. (It would be preferable to keep the sites fixed at a given geometry i.e. the reference minimum used in fitting the potentials, but this was not possible in DL_POLY as the shells must be allowed to move.) It was found that the agreement between total energy of the dimer in simulation and total energy of the dimer using point polarisabilities in Orient did not necessarily improve as the charge size increased, see Table 5.1. To avoid delving into this complexity, a charge of $q_s = 1$ a.u. was used, though a charge somewhere between 1.0 and 2.0 may have been preferable.

Setting the interactions to occur between cores and not shells improved the match between the point-polarisable model and the shell model, giving a minimum energy of -20.983 kJ mol$^{-1}$ for $q_s = 1$ (compared with the reference value of -20.609 kJ mol$^{-1}$). Unfortunately, DL_POLY at that time did not allow for a core-core approach in large-scale simulations (i.e. bulk water or ice).

| $q_s$ | Energy / kJ mol$^{-1}$ |
|---|---|
| 0.5 | −19.012 |
| 1.0 | −19.436 |
| 1.5 | −19.700 |
| 2.0 | −17.573 |
| (point polarisable) | −20.609 |

Table 5.1: Energies of the minimum configuration reminimised for different shell charges and for the point-polarisable model they are intended to reproduce.

## 5.3 Estimating the Diffusion Coefficient

When discussing the motion of particles from an area of high concentration into lower concentration, the diffusion coefficient $D$ is defined by Fick's law,

$$\mathbf{j} = -D\nabla c \tag{5.5}$$

It relates the molar flux $\mathbf{j}$ to the gradient of the concentration $c$. In self-diffusion the particles are labelled but moving within a material of identical particles, in this sense it describes the tendency of individual molecules to travel within a bulk fluid. When moving in three dimensions, the self-diffusion is related to the mean-square displacement $\langle r(t)^2 \rangle$ by the relation

$$\partial_t \langle r(t)^2 \rangle = 6D \tag{5.6}$$

To account for periodic boundary conditions, the displacement $r(t)$ should not be thought of as the measured distance from some start point but instead should be the time-integral of the velocity.

It is established that estimates of the diffusion coefficient by MD simulation depend on the size of the system: In particular, the diffusion coefficient is underestimated in a cubic unit cell by factor proportional to $1/L$ where $L$ is the width of the cell, or correspondingly by $N^{-1/3}$.

For this reason, a system of 6144 molecules was used, this is large compared with, for example, the 256 molecules used in finding the diffusion coefficient of MB-Pol [146].

All simulations in this section used constant temperature and pressure.

### 5.3.1 Calculating the diffusion coefficient for water models

For water, the cell was prepared by melting a cell of ice Ih $8\times12\times8$ unit cells in size at 400 K and 1 atm using the $+\Delta Q$ charge model. This was then allowed to run for 160 ps at the desired temperature and water model of interest before data collection began. For the polarisable model, the first 10ps were run with a timestep of 0.2 fs, allowing the shells to position themselves; and the remainder of the simulation ran with a 1 fs timestep. All other simulations used a 2 fs timestep.

Simulations were run for a minimum of 400ps, however the MSD does not grow linearly with time across the entire region, this was particularly noticeable for the $+\Delta Q$ model, see Fig 5.1. Because of this, the MSD was calculated across the following regions: 25 to 250 ps for the polarisable model; 300 to 500 ps for the $+\Delta Q/2$ model; and just 40 to 90ps for the $+\Delta Q$ model. The results are discussed in Section 4.5.1.

### 5.3.2 Calculating the diffusion coefficient for methane models

For the methane models the cell was prepared using a regular grid of molecules. Two reference points were chosen from available experimental data, 303.3 K and 333.1 K at 296.1 atm, these particular data were chosen because it was very close to the intended clathrate simulation conditions. After equilibrating, simulations were run for a minimum of 120 ps. All methane models gave a plot very close to a straight line, so no consideration was taken with regards to the region used to calculate the gradient. The results are discussed in Section 4.5.2.

## 5.4 Estimating the Melting Temperature of a Water Model

To estimate the melting temperature the intent was to use the same method as Fernández et al. (2006) [142]. This involves running MD simulations of a water/ice interface at 1 atmosphere of pressure and a range of chosen temperatures. If the temperature is above the melting point, the ice will begin to melt. To maintain the temperature, the thermostat must add energy to the system and so the total energy increases (similarly if the liquid water freezes, the total energy must decrease). Fernández et al. are able to ascertain the melting temperature to an accuracy of 2° using this method. Using the same methodology, though, it was found that the total energy changed very slowly and that in general if the block of ice was melting it could be easily seen visually. An example of this is given in Figures 5.2(a) ad 5.2(b). To estimate the melting temperature the simulation ran for up to 2 ns, from this there is generally very little change in the total system energy close to the melting temperature but an upper limit can be placed if melting is apparent.

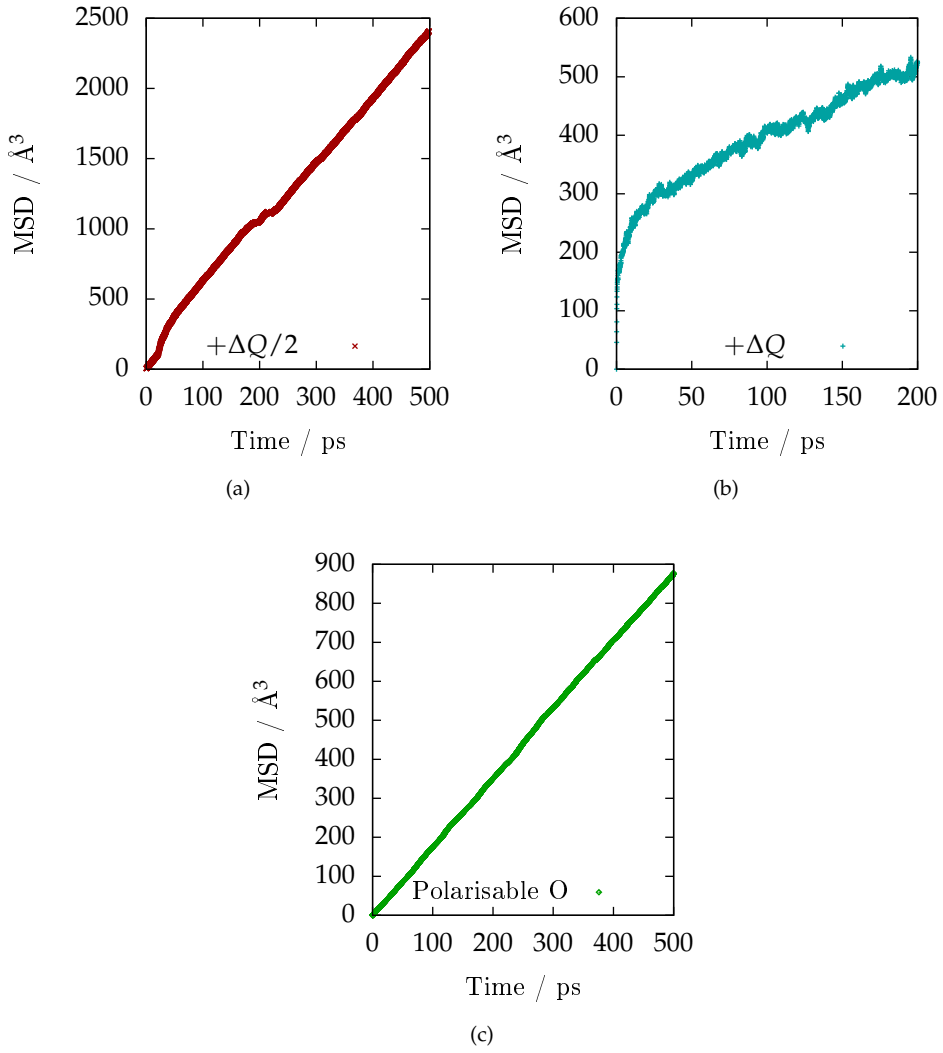
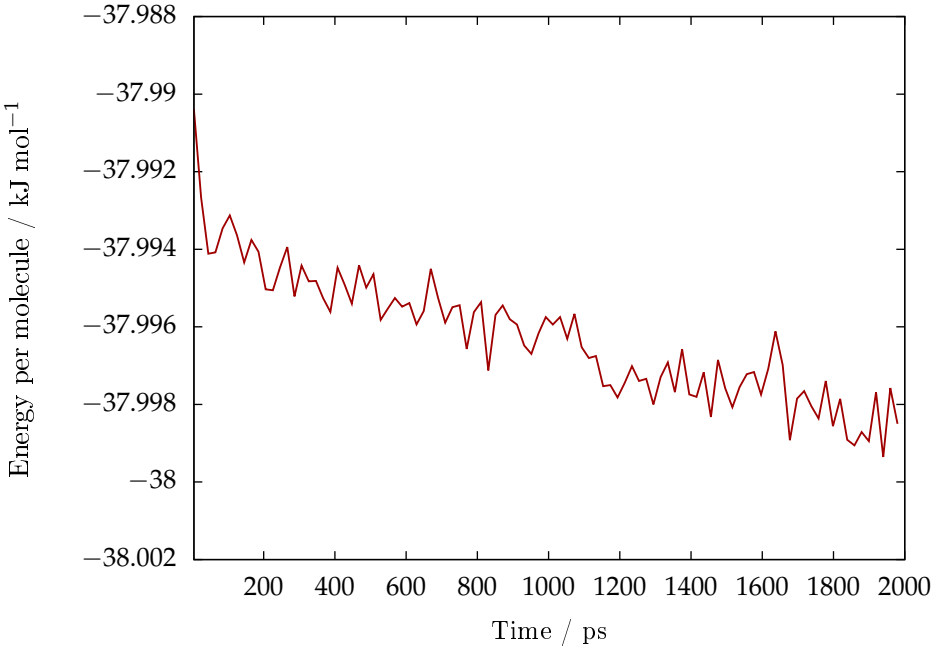
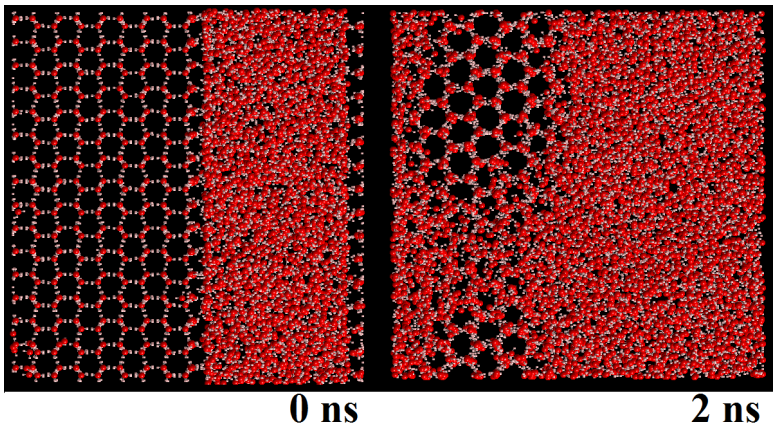



Figure 5.1: Plots of mean-square displacement against time for: (a) the $+\Delta Q/2$ charge model, (b) the $+\Delta Q$ charge model, (c) the polarisable oxygen model of water. These plots were used to find diffusion coefficients.




(a)




(b)

Figure 5.2: 5.2(a) Example plot of energy per molecule against time at 170 K and 1 atm for the $+\Delta Q/2$ model, this is expected to increase as the ice melts. 5.2(b) Cell at the start of the simulation and after 2 nanoseconds, showing that the ice has melted.

## 5.5 Simulation of Clathrate Melting

When methane clathrate is melted the methane-water solution is supersaturated and so the methane will generally be removed form the water by forming bubbles on the nanometre scale. This is sometimes associated with the clathrate "memory effect", which is strongest when the clathrate is brought under conditions to reform soon after melting, but only provided the temperature is within a few degrees of the melting point [152, 153].

One suggested explanation for the memory effect is that shortly after melting the methane molecules might exist for some time in something close to an ordered structure, the "guest supersaturation" hypothesis. A conflicting explanation is that the surface of methane nanobubbles serve as the nucleation points [76].

For local ordering to result in nucleation, there must be methane molecules dissolved in liquid water at a comparable density to in clathrate. Although simulations are unlikely to fall within just a few degrees of the clathrate melting temperature, which is obviously model-dependent and has not been determined, it is relatively straight forward and not too computationally demanding to study decomposition of methane clathrate in cases where the methane then joins into a bubble. There are still relatively few studies of this nature, generally these use the TIP4P/Ice and UAM models [72, 96].

The formation of methane bubbles was studied through MD simulation as follows: Starting with a simulation cell $4 \times 4 \times 12$ unit cells in size, a region of methane clathrate was melted under NVT conditions at 450 K and starting pressure of 300 atm until a methane bubble had formed, with all other sites kept "frozen" in place on either side. The simulation cell was then allowed to expand to its natural volume under NPT conditions at 300 atm and a chosen temperature. Following this, the simulation was run under NVT conditions until all layers of the clathrate cell had melted. The number of methane molecules in each twelfth of the cell (i.e. roughly each layer of the clathrate, or containing 128 guests on average) is then plotted against time, from which the transport of methane into the bubble can be understood. In discussing these simulations, we will assume that the bubble spans regions 1,2, 11 and 12, i.e. the outsides of the cell, see Fig 5.3.

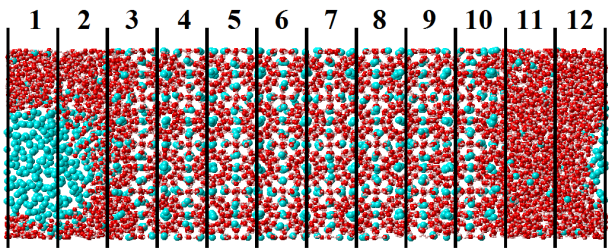



Figure 5.3: The regions used in studying the transport of methane as the clathrate melts. At the start of the simulation, regions 1,2,11 and 12 are melted.

Three combinations of water and methane models were used: The TIP4P/Ice model and the United Atom Methane (UAM) model; the $+\Delta Q$ model and the 5 site methane model; and the $+\Delta Q$ and UAM model. In the interest of simplicity, the $+\Delta Q$-UAM interaction used the same potential as the TIP4P/Ice-UAM interaction.

Each combination of models was used at two temperatures, one giving "fast melting" i.e.

the entire cell melting in ∼1ns and one giving "slow melting" lasting ∼10 ns. For simulations using TIP4P/Ice the fast melting happened at 380 K and 300 atm, and slow melting at 350 K and 300 atm. The $+\Delta Q$ model has a lower melting temperature than TIP4P/Ice and used 350 K for fast melting and 320 K for slow melting.

Using the UAM model and either water model gave a qualitatively similar melting, and these models are discussed together. Using the 5 site methane model gave quite different melting and this will be discussed afterwards.

### 5.5.1 Slow melting using the UAM model

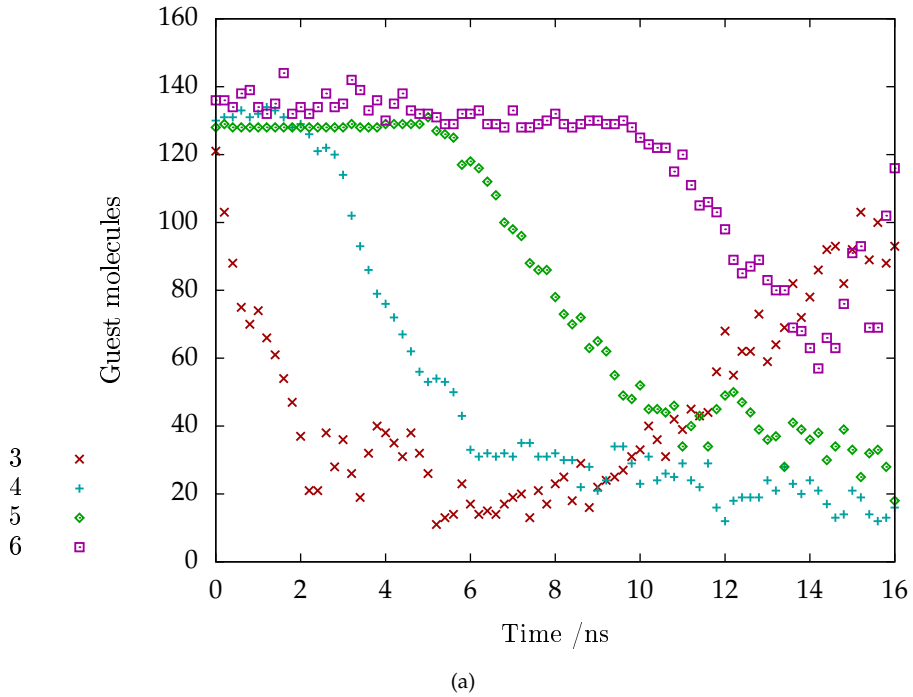
Under slow melting, the guest molecules can be seen leaving each region as the layer melts. The methane in each region begins decreasing as soon as melting begins, for example region 5 begins to empty as soon as region 4 is melted. Snapshots are given just before each region begins to empty which illustrate this. Regions next to an existing bubble have as few as 10 guests each, an order of magnitude smaller than at the start of the simulation. Further from this the regions begins to empty but this emptying slows when the next region melts, so that the number of methane molecules in each successive layer seems to level off at a higher value. After three or four layers have melted, new methane bubbles begin to form.

#### 5.5.1.1 TIP4P/Ice model details

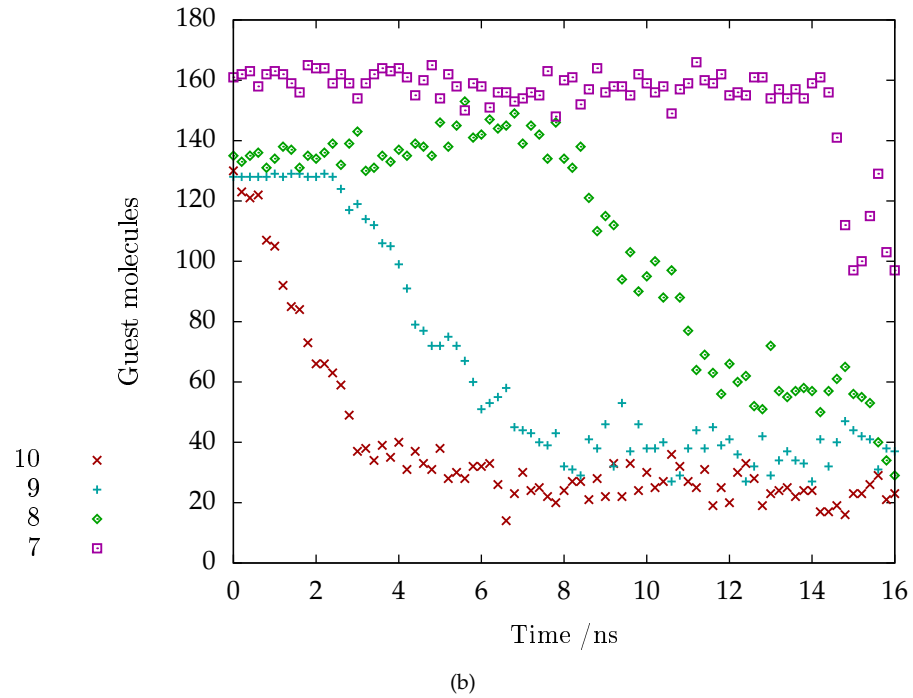
Melting eight layers of TIP4P/Ice and UAM clathrate at 350 K and 300 atm took around 15 nanoseconds. Methane in regions 5 and 8 (i.e. next to the final layers) fell to around 40 guest molecules until the final layers melted to form a new bubble, see Figures 5.4 and 5.5. It should be noted the final layer is being melted from both sides so it may behave differently. Regions which do not include part of a bubble have between 15 and 40 guest molecules.

#### 5.5.1.2 $+\Delta Q$ model details

To melt eight layers using the $+\Delta Q$ model and UAM at 320 K and 300 atm took around 8 nanoseconds. There are two new bubbles; one which formed at around 5.8ns in regions 8 and 9 and one which formed around the final layer. In regions without bubbles, the number of methane molecules has fallen to between 10 and 40. See Figures 5.6 and 5.7

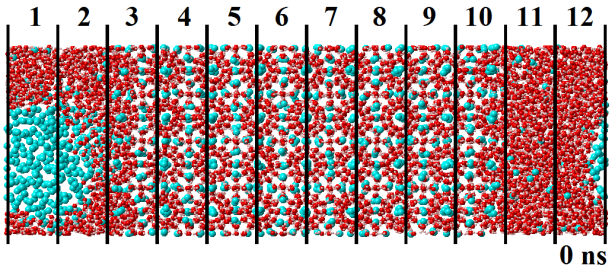



(a)

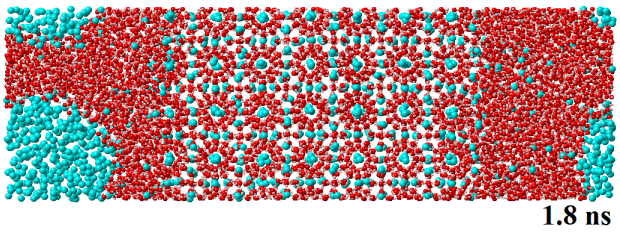



(b)

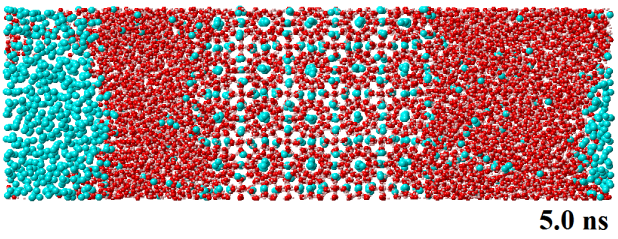
Figure 5.4: Plots of the number of methane molecules in each region as the clathrate melts at 350 K and 300 atm, using the TIP4P/Ice and UAM models. At the start of the simulation, regions 1,2,11 and 12 are already melted and contain a methane bubble. The final regions to melt are 6 and 7, about which a second bubble forms. The increase in region 3 is due to the starting bubble growing into the region. Each region begins to lose methane as soon as the region beside it has melted, see also Fig 5.5.



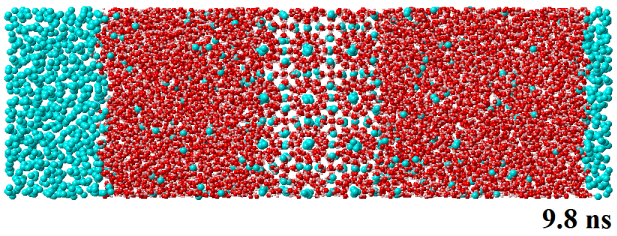

(a)



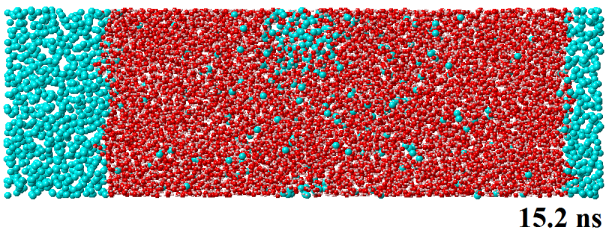

(b)




(c)

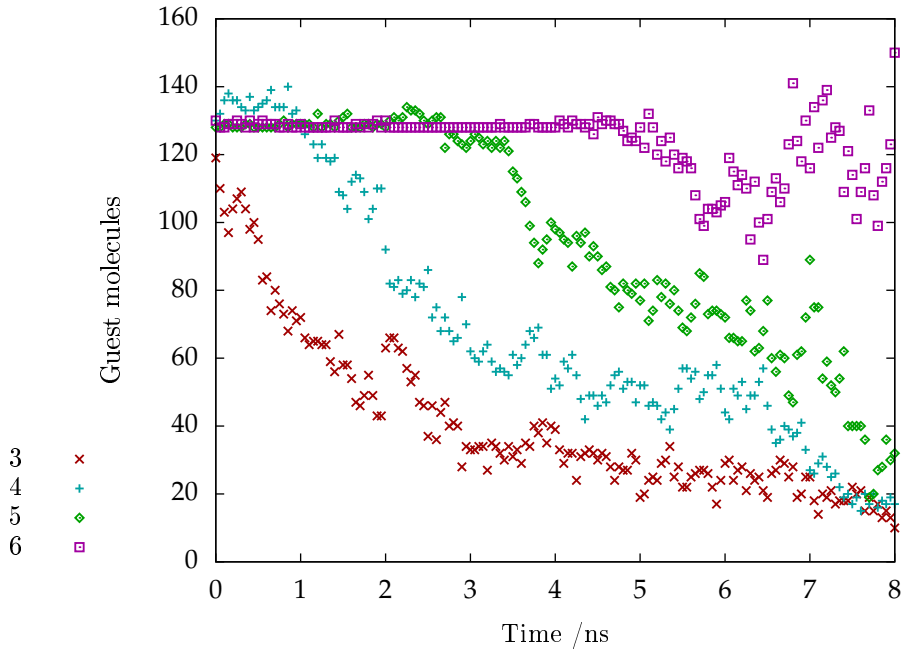



(d)

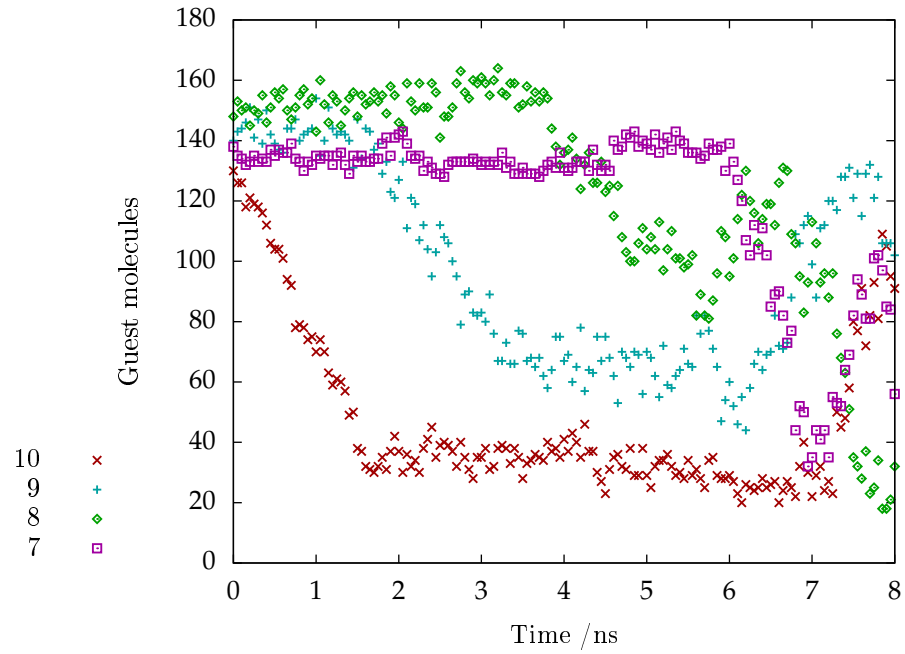



(e)

Figure 5.5: Melting each layer of clathrate in regions 4, 5, and 6, shown as melting in each layer completes using the TIP4P/Ice and UAM models. Melting finishes just before guest molecules begin escaping each region, see Fig 5.6(b). As the final layer melts, a second methane bubble forms (5.5(e)). Oxygen in red, single-site methane in blue.

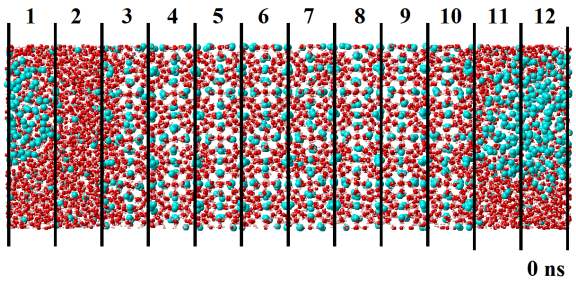



(a)

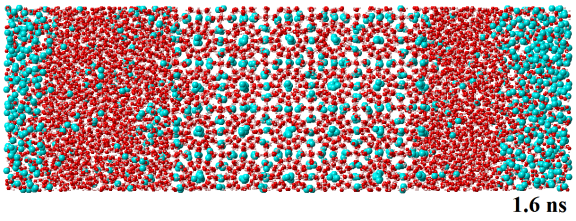



(b)

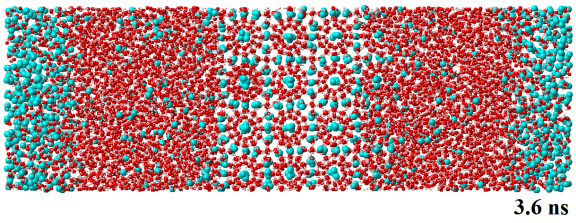
Figure 5.6: Plots of the number of methane molecules in each region as the clathrate melts at 320 K and 300 atm, using the $+\Delta Q$ and UAM models. At the start of the simulation, regions 1,2,11 and 12 are already melted and contain a methane bubble. The formation of a new bubbles can be seen in the peaks at at region 8 at 6.4 nanoseconds and region 6 at around 7.1 nanoseconds. Each region begins to lose methane as soon as the region beside it has melted, see also Fig 5.7.



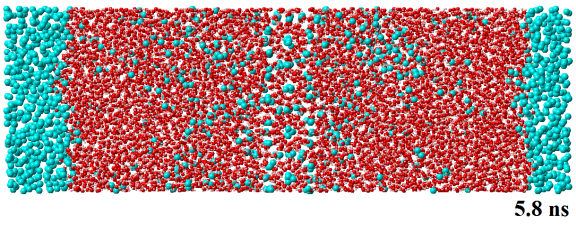

(a)



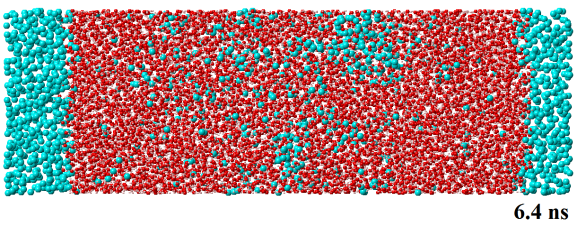

(b)




(c)




(d)




(e)

Figure 5.7: Melting each layer of clathrate in regions 10, 9, and 8 shown as melting in each layer completes using the $+\Delta Q$ and UAM models. Melting finishes just before guest molecules begin escaping each region, see Fig 5.6(b). At 6.4 ns (5.7(e)) a new bubble can be seen in region 8, corresponding to the peak in number of guest molecules seen in Fig 5.6(b). Oxygen atoms are shown in red, single-site methane in blue.

### 5.5.2 Fast melting using the UAM model

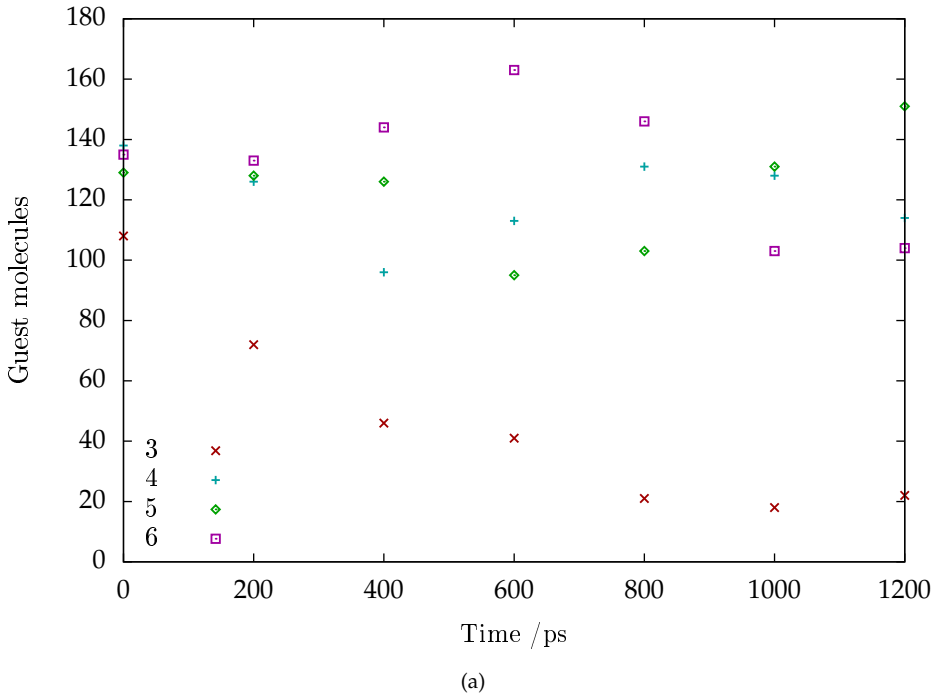
Under fast melting each layer melts before the number of methane molecules in the previous region has significantly decreased, for example region 4 will melt before the methane in region 3 can escape towards the bubble. Because of this methane may not leave the region and can begin to form new bubbles before region 4 has melted. By the time all layers have melted, only one of twenty four regions using either water model contained fewer than 60 guest molecules.

#### 5.5.2.1 TIP4P/Ice model details

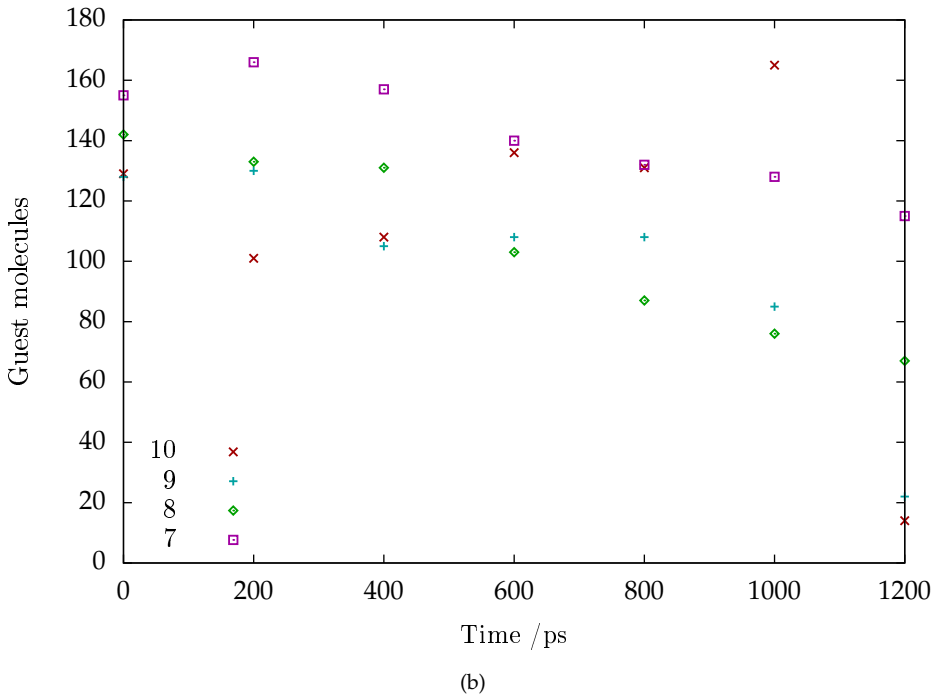
Melting eight layers of TIP4P/Ice and UAM clathrate at 380 K and 300 atm was completed by 1 ns. The first bubble formed in region 10 at around 400 ps and a second had formed in region 3 by 1ns. During the first nanosecond all melting regions except region 3 had 75 methane molecules or more, 200 ps after melting region 3 and two other regions were at about 20 methane molecules each.

#### 5.5.2.2 $+\Delta Q$ model details

Melting eight layers using the $+\Delta Q$ model and UAM at 350 K and 300 atm took around 0.7 nanoseconds. Two new bubbles had formed across regions 2 and 3 by 350 ns shown in Fig 5.11(b). By 500ps these had joined together and a new bubble had formed in region 9, these remained until the end of the simulation (Figs 5.11(d) and 5.11(d)). Throughout the simulation, no region had fewer than 60 methane molecules, as shown in Fig 5.10.

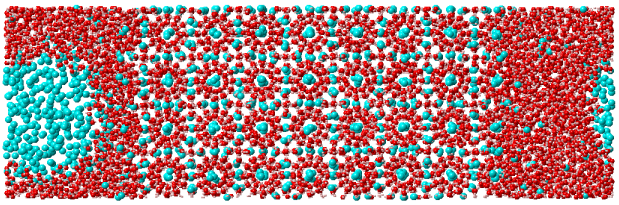



(a)

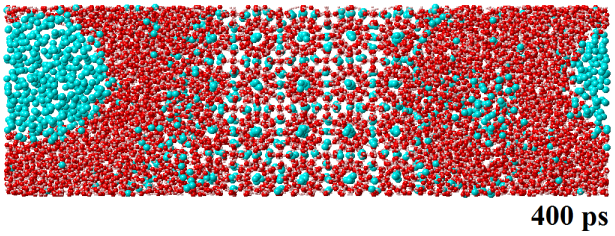



(b)

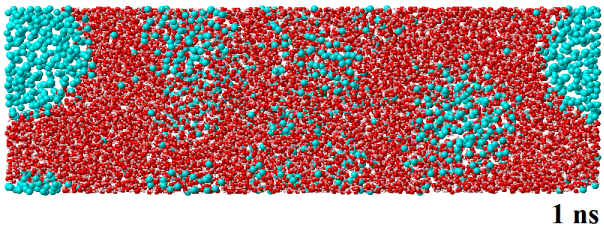
Figure 5.8: Plots of the number of methane molecules in each region as the clathrate melts at 380 K and 300 atm, using the TIP4P/Ice and UAM models. At the start of the simulation, regions 1,2,11 and 12 are already melted and contain a methane bubble. All cells are melted by 1 ns as shown in Fig 5.9, but in this time only region 3 (closest to the starting bubble) has experienced a large decrease in the number of guest molecules.




(a)

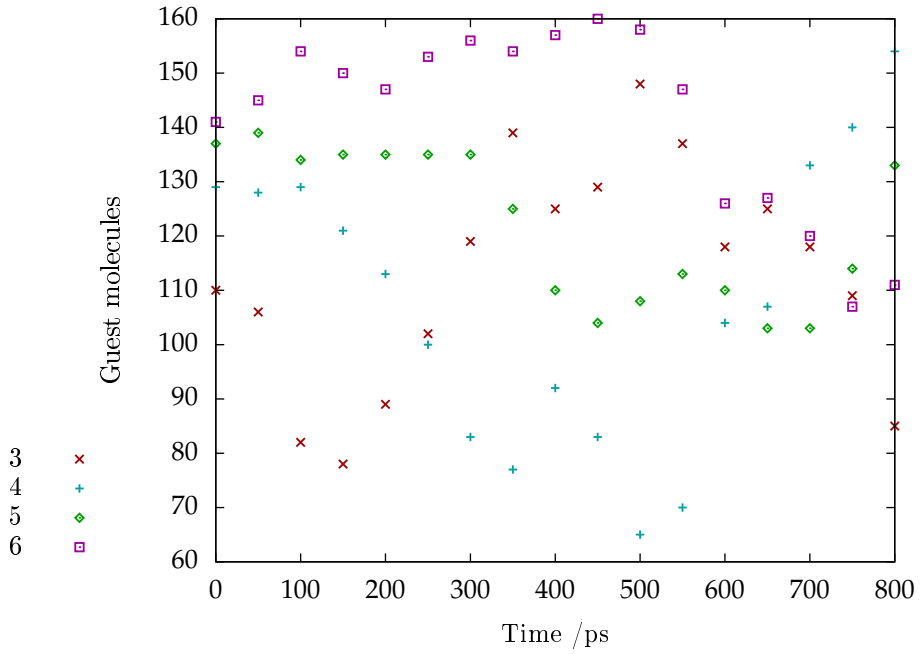



(b)

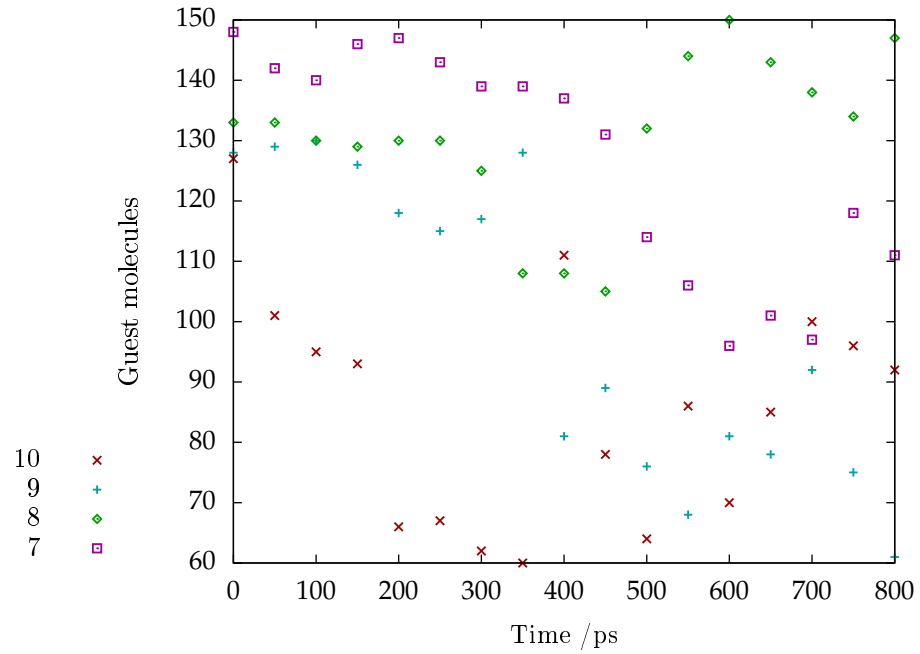



(c)

Figure 5.9: Formation of bubbles melting at 380K and 300 atm, using the TIP4P/Ice and UAM models. A bubble is formed at 400ps and continues to grow. By 1ns a second bubble has formed. Within this time, there is relatively little flow of methane from melted regions, as shown in Fig 5.8. Oxygen atoms are shown in red, single-site methane in blue.

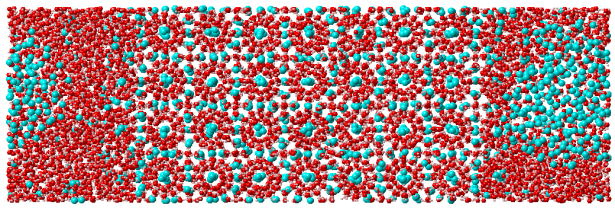



(a)



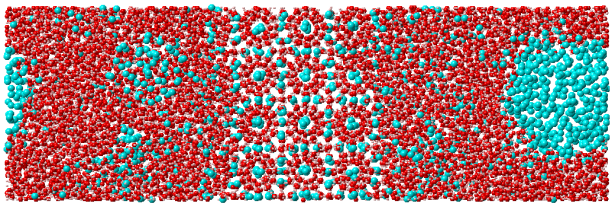

(b)

Figure 5.10: Plots of the number of methane molecules in each region as the clathrate melts at 350 K and 300 atm, using the $+\Delta Q$ and UAM models. At the start of the simulation, regions 1,2,11 and 12 are already melted and contain a methane bubble. The number of methane molecules in regions 4, 9 and 10 falls to around half its initial value before increasing again, this is because of the movement of the newly formed bubbles seen in Fig 5.11.



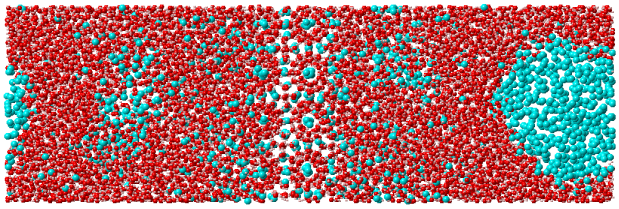
0 ps

(a)



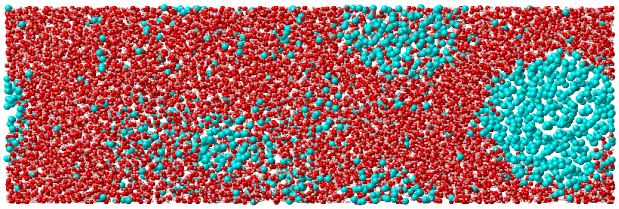
350 ps

(b)



500 ps

(c)



700 ps

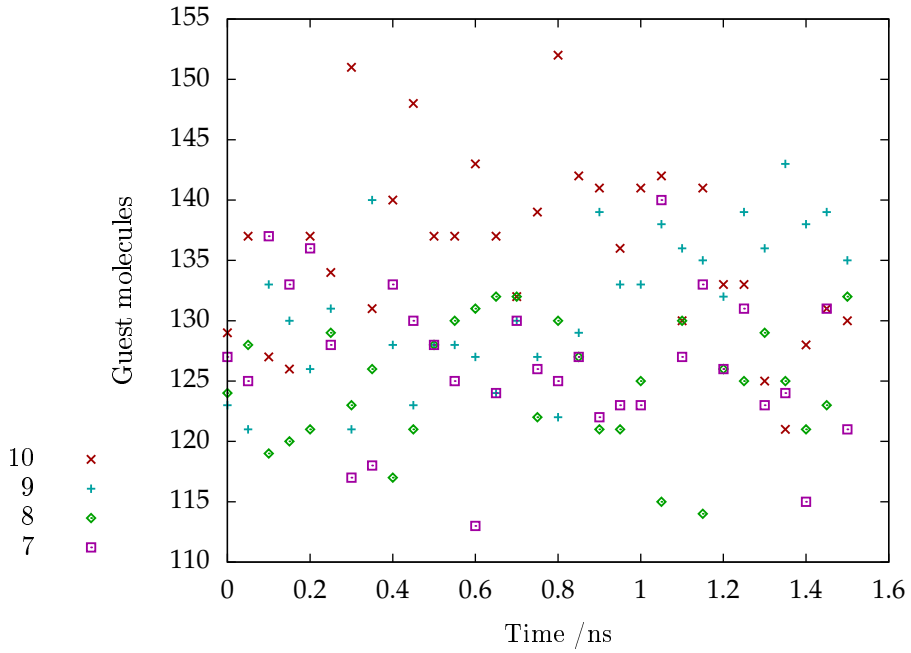
(d)

Figure 5.11: Melting the clathrate at 350 K and 300 atm, using the $+\Delta Q$ and UAM models. As the clathrate is melted, the methane tends to form new bubbles rather than leave a region, see Fig 5.10. Oxygen atoms are shown in red, single-site methane in blue.

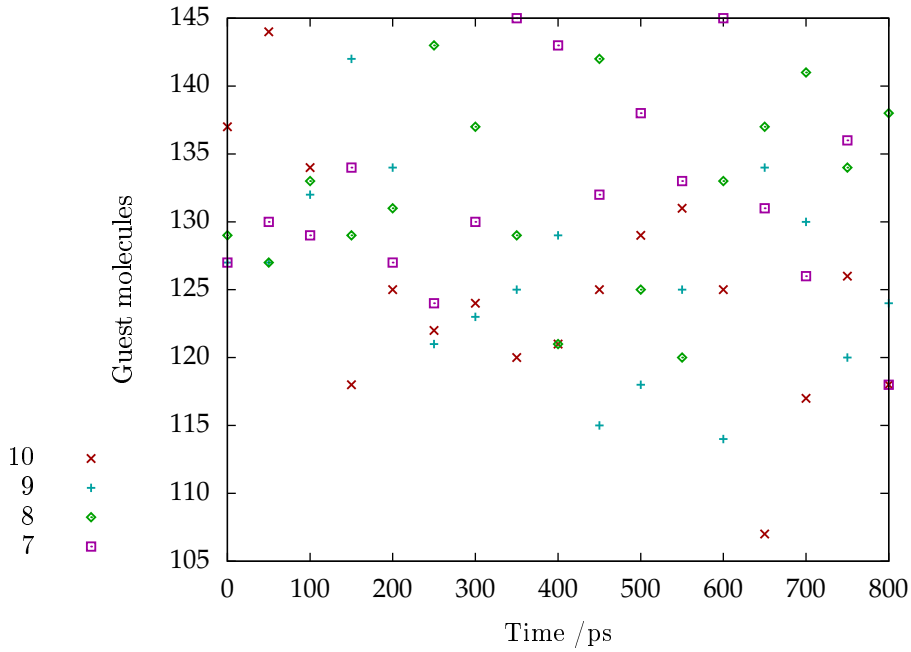
### 5.5.3 Melting using the five-site methane model

Using the five-site methane model and the $+\Delta Q$ water model, the melting rate was similar to using the $+\Delta Q$ and UAM models: At 350 K and 300 atm, all layers melted in about 0.6ns compared with 0.7 ns using the UAM model. At 320 K and 300 atm, the first layer melted after around 1.7 ns compared with 1.6 ns, and the second melted at around 3.0 ns compared with 3.3 ns, after 4ns the simulation was ended.

Unlike with the UAM model, the methane did not form into bubbles at either temperature, or during the preparation period at 450 K. In fact the amount of methane in each region stayed relatively constant throughout the simulations. The methane count is illustrated for one half of the cell at both temperatures in Fig 5.12, and the lack of bubbles is also clearly demonstrated in simulation snapshots at either temperature in Fig 5.13.

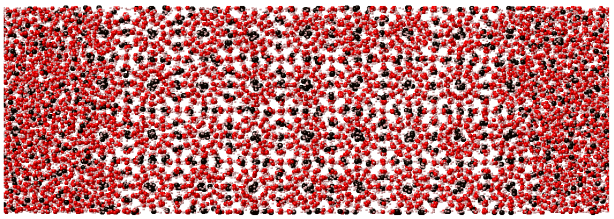



(a)



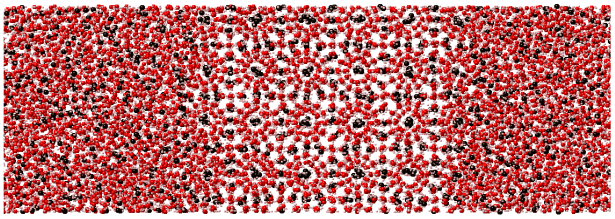

(b)

Figure 5.12: Plots of the number of methane molecules in each region as the clathrate melts at 300 atm and at (a) 320 K and (b) 350 K, using the $+\Delta Q$ and 5-site methane models. There is very little change in the number of methane molecules in any region, even after layers have melted.



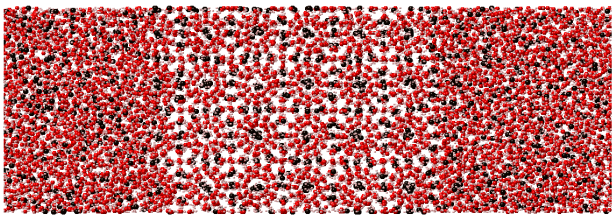
0 ns

(a)



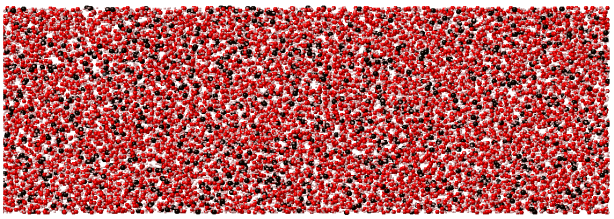
4 ns

(b)



0 ps

(c)



700 ps

(d)

Figure 5.13: Melting the clathrate at 300 atm and at 320 K ((a) and (b)) and at 350 K ((c) and (d)), using the $+\Delta Q$ and five-site methane models. At either temperature, there is no formation of methane bubbles. Oxygen atoms are shown in red, carbon atoms in black.

## 5.6 Additional clathrate simulations

### 5.6.1 Clathrate-forming conditions

Simulations were carried out for each model combination using constant pressure and temperature under known clathrate-forming conditions of 280 K and 200 atm (i.e. circa 2 km ocean depth). These were carried out firstly in a continuous clathrate cell for 100ps, so that the difference in density could be compared between models. Following this these conditions were used with the same cell as the previous section, i.e. 4 by 4 by 12 unit cells with a 4 by 4 by 4 section melted. These simulations ran for 4 ns.

Using either the TIP4P/Ice and the UAM models or the $+\Delta Q$ and five-site methane models there were some signs of clathrate growth under these conditions. There was no visible growth using the $+\Delta Q$ and UAM models, see Fig 5.14. The five-site methane would be expected to give faster growth due to the availability of dissolved methane.

The densities of the clathrate using different models under these conditions are given in Table 5.2. The $+\Delta Q$ model gives very similar densities of 0.85 g cm$^{-3}$ using UAM methane or 0.86 g cm$^{-3}$ using the five-site methane model, whereas using the UAM and TIP4P/Ice models gives a slightly higher density of 0.92 g cm$^{-3}$. There is a lack of experimental data for sI methane clathrate at these conditions, however estimates of the density at 100 bar (98.7 atm) and 280K give a density between 0.90 and 0.92 g cm$^{-3}$ [154].

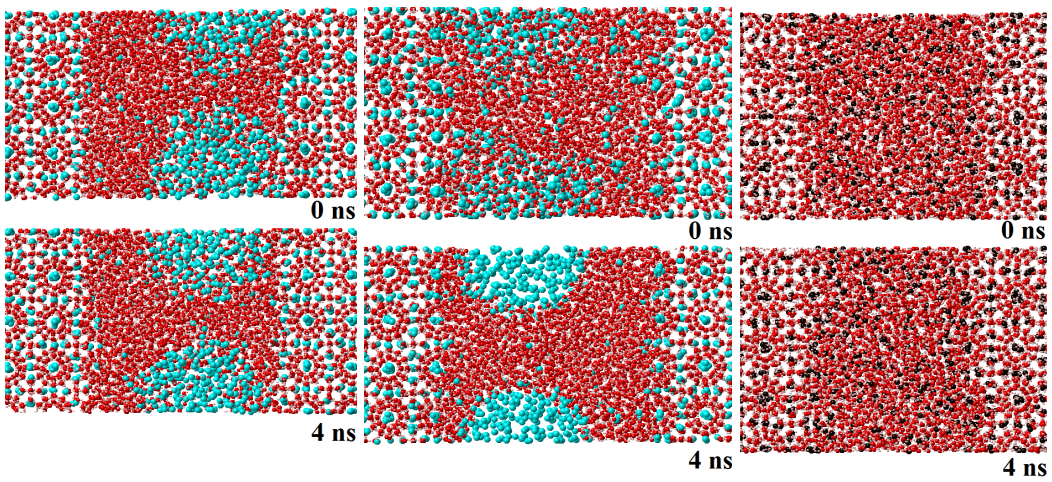



Figure 5.14: A clathrate interface at 280 K and 200 atm for the UAM and TIP4P/Ice models (5.14(a)), the UAM and $+\Delta Q$ models (5.14(b)), and the $+\Delta Q$ and five site methane models (5.14(c)). The interface is shown at the start of the simulation and after 4 ns. There are signs of clathrate growth using TIP4P/Ice and the UAM model, and using the $+\Delta Q$ and the five-site methane model, but there is no sign of growth using the $+\Delta Q$ and UAM models. Oxygen atoms are shown in red, single site methane in blue and carbon atoms in black.

### 5.6.2 Simulations under ambient conditions

For all previous conditions the five-site methane model had not led to bubble formation, in contrast to previous simulations. To better evaluate the five-site methane model, NPT

| | Density / g cm$^{-3}$ |
|---|---|
| TIP4P/Ice and UAM | 0.9201 |
| $+\Delta Q$ and UAM | 0.8466 |
| $+\Delta Q$ and 5site methane | 0.8581 |

Table 5.2: Densities of methane clathrate using different models at 280 K and 200 atm.

simulations were also carried out at ambient conditions, i.e. 298.15 K and 1 atm, under which it is well-established that methane does not dissolve in water.

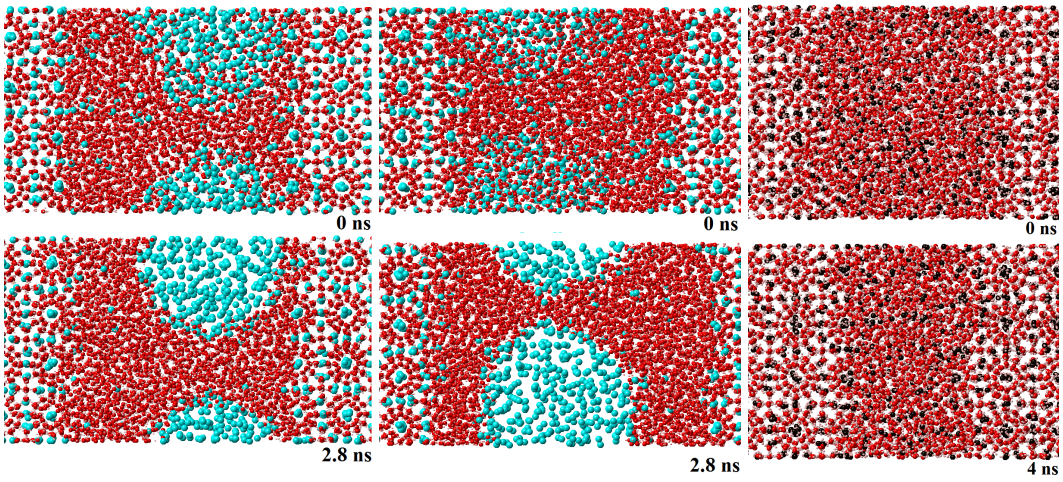
#### 5.6.2.1 Starting from an an interface

At fist the same setup was used as in the previous interface simulations. The $+\Delta Q$ and TIP4P/Ice models both showed the methane within the melted region run into the bubble after 2.8 ns. When the $+\Delta Q$ and five-site methane models were used; not only did a bubble not form; but the clathrate began to reform, as shown after 4 ns in Fig 5.15. This is particularly surprising as the melting point for ice Ih using the $+\Delta Q$ model at 1 atm is at least 50 degrees below this temperature (Sec. 4.4).

#### 5.6.2.2 Starting from a methane bubble in water

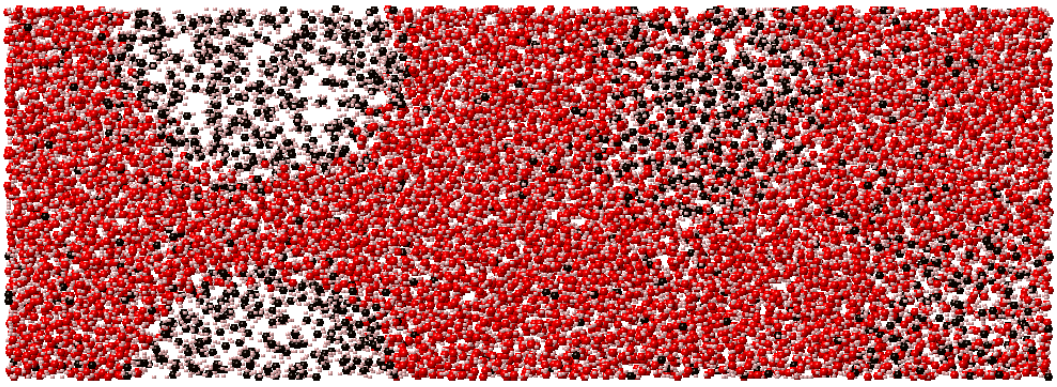
Next these temperature and pressure conditions were used alongside a starting configuration taken from the complete melting of the clathrate at 350 K and 300 atm. In this case it was found that no bubble formation took place within 1ns. Finally, a starting configuration was produced which contained three methane bubbles. To do this, the charges on the methane model were made very small (-0.04 for the carbon atom and 0.01 for the hydrogen atoms). Other methane-methane and methane-water interactions were replaced with the UAM and TIP4P/Ice interactions, except that the methane-methane well depth was increased from 1.2301 kJ mol$^{-1}$ to 2 kJ mol$^{-1}$ to encourage bubbles to form. After 600 ns three bubbles had formed as shown in Fig 5.16(a). Following this preparation stage, the potentials were returned to their original form.

The simulation ran for 2 ns under NVT conditions, after which the largest bubble survived. The two smallest bubbles have not been swallowed by the larger bubble but instead have dissolved. It might be ideal to try and simulate under constant pressure and temperature, but the low density of methane gas at this temperature and pressure makes the simulation impractical. Such a simulation was able to run for about 750 ps, during which time the same bubble was seen to survive (and expand) while the others dissolved. Snapshots from both simulations are shown in Figure 5.16

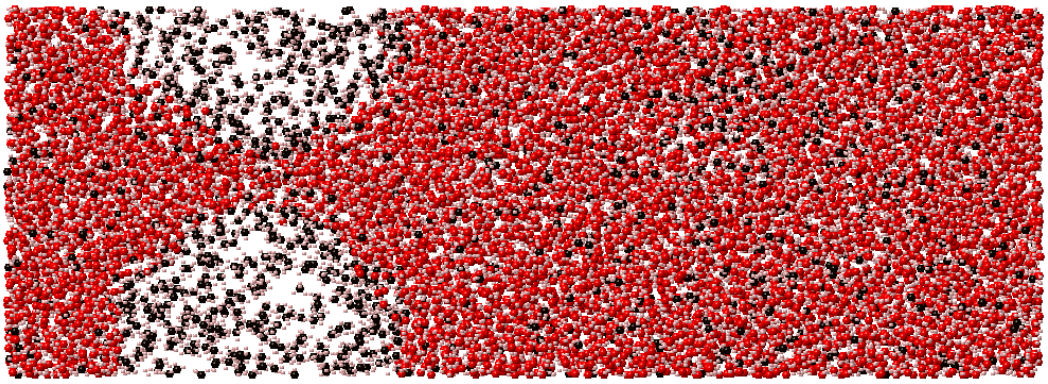



(a) (b) (c)

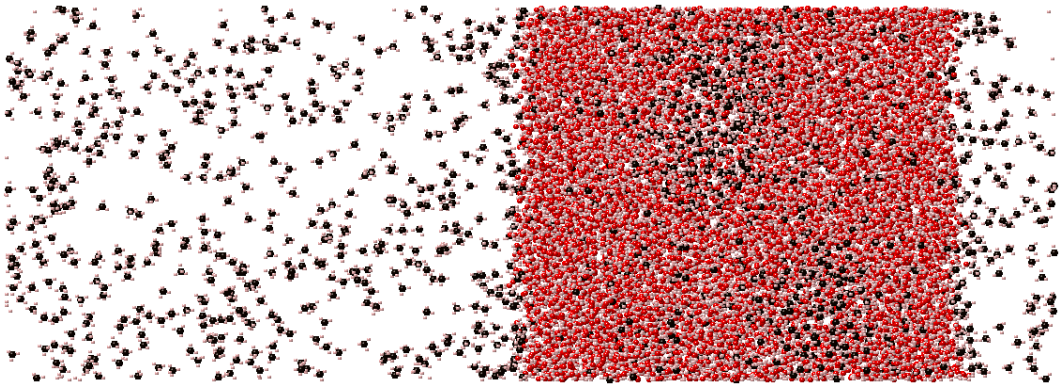
Figure 5.15: A clathrate interface at 298.15 K and 1 atm for the UAM and TIP4P/Ice models (5.15(a)), the UAM and $+\Delta Q$ models (5.15(b)), and the $+\Delta Q$ and five site methane models (5.15(c)). In the first two cases the bubble has grown and the region around it seems to have cleared, but when using the five-site methane model no bubble develops and the clathrate reforms. Oxygen atoms are shown in red, single site methane in blue and carbon atoms in black.



(a)



(b)



(c)

Figure 5.16: The starting configuration (*a*) and the configuration after: (*b*) 2 ns under NVT conditions at 293.81 K. (*c*) 600 ps under NPT conditions at 293.81 K and 1 atm. In both cases the larger bubbles has survived and the smaller bubbles have dissolved. Oxygen atoms are shown in red, carbon atoms in black.

## 5.7 Discussion of clathrate simulations

The clathrate simulations were for two purposes; to test that the potentials developed in Chapter 3 did not behave in any peculiar way when used for studying clathrate dynamics and to help understand the flow of methane under clathrate decomposition.

With regards to the first point, it is clear that clathrate simulated using the five-site methane model and accompanying water-methane potential developed in Chapter 3 behaves qualitatively differently from clathrate using the UAM model of methane. Melting the former gave methane molecules dissolved in water, and at a clathrate interface under ambient conditions these loose molecules apparently facilitated clathrate growth.

As clathrates have generally been found to have similar melting points to ice, it is especially notable that the melting temperature of ice Ih using the $+\Delta Q$ model is 240 K or lower but that the clathrate continued to grow at 298.15 K. This at least highlights the importance of guest availability in the formation of sI clathrate.

The simulations on decomposition using the UAM methane model did not seem to vary much between the TIP4P/Ice and $+\Delta Q$ water models, except that the melting rate was faster with the $+\Delta Q$ model, as would be expected from its lower melting temperature for ice Ih. The general view of methane flow during sI clathrate decomposition near a bubble is as follows: At low temperatures, the guest molecules will escape as the clathrate melts and will flow into the bubble, so that the number of methane molecules in one layer can fall by an order of magnitude by the time the next layer has began to melt. It will then stay close to constant as it is fed by guests escaping the new layer. At higher temperatures several layers of clathrate may melt before methane has had time to flow into the bubble, and instead new bubbles are formed.

As a further detail, when melting at low temperatures each successive layer seems to level off at a slightly higher number of methane molecules. In the simulation where melting was slowest, no methane bubbles formed until the final layer melted, which involves melting from both sides and is perhaps not representative of normal clathrate decomposition. Based on these simulations it is not inconceivable that when decomposition happens within a few degrees of the melting point there is still less inclination for bubbles to form, and that clathrate growth is encouraged by the presence of dissolved methane. Further, the experimental evidence to date suggests that while nucleation happens more readily during clathrate reformation, the rate of growth following nucleation is not subject to any "memory effect" [153]. For residual dissolved methane to account for this behaviour, it would need to survive only in small regions which would serve as nucleation points. These simulations do not show nucleation can happen under such circumstances or allow comment on its likelihood relative to nucleation about a gas bubble, say, but do suggest that these initial conditions are possible.

### 5.7.1 Further discussion on methane-water potentials

The water-methane potential which accompanied the five-site model was designed to work with the isotropic water model with three polarisable sites, developed in Section 3.2. The water model used in simulations is non-polarisable but has an electrostatic model designed to account for this in bulk water-water interactions. This leads to overbinding for an individual water dimer.

The methane-water interaction also tends to be slightly too negative, as shown in Fig 5.17. For a single methane molecule surrounded by water molecules, as in a clathrate, the $+\Delta Q$ model of water might be an appropriate description (as far as non-polarisable models allow) since the water will still be polarised by its neighbours. In cases where there are methane molecules interacting with one another as well as water molecules, for instance a water molecule buried among early stages of a methane nanobubble, then the water-methane interaction from the model is likely to be stronger than it should. This may discourage the formation or stability of methane nanobubbles. Although the problem lies in the interaction of charge models, it may be possible to rescale the exchange-repulsion part of the interaction to give a more accurate methane-water potential.

The UAM model and TIP4P/Ice models were used together and their interaction was made according to the Lorentz-Berthelot combining rules, i.e. for Lennard-Jones model

$$E_{LJ}^{Me} = 4\epsilon^{Me}\left((\frac{\sigma^{Me}}{R})^{12} - (\frac{\sigma^{Me}}{R})^{6}\right) \tag{5.7}$$

for methane and a similar term $E_{LJ}^{Wa}$ for water, the interaction can be approximated by

$$\epsilon = (\epsilon^{Me}\epsilon^{Wa})^{1/2}\sigma = (\sigma^{Me} + \sigma^{Wa})/2 \tag{5.8}$$

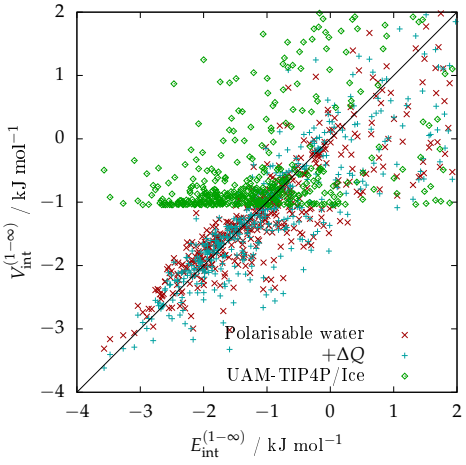
This approach is found often in clathrate decomposition and nucleation literature, for example nucleation work by Walsh et al. [83, 87, 155] or decomposition work by Baherzedadeh et al. [72, 96]. However the use of this combining rule cannot result in the correct combination of the two potentials: The UAM Lennard-Jones terms are used to describe the total energy of the molecule, whereas the TIP4P/Ice model consists of both a Lennard-Jones site and a set of charges. In effect, by using combining rules for a Lennard-Jones potential, the methane is only getting part of the interaction energy with the water molecule. The underbinding of the methane-water potential from mixing rules was discussed in 2005 by Konrad and Lankau [156], and in 2006 by Docherty et al. [157].

In the case of TIP4P/Ice, the term $\epsilon$ is actually only around 0.88 kJ mol$^{-1}$. This is lower than the UAM model's term of 1.23 kJ mol$^{-1}$, meaning that the water-methane interaction (1.04 kJ mol$^{-1}$) is actually weaker than the methane-methane interaction when using this method. See Fig 5.17. Such a weak methane-water interaction must mean methane can be transported through water with very little hindrance, the end result is possibly unrealistically fast gathering of methane into nanobubbles upon clathrate melting.

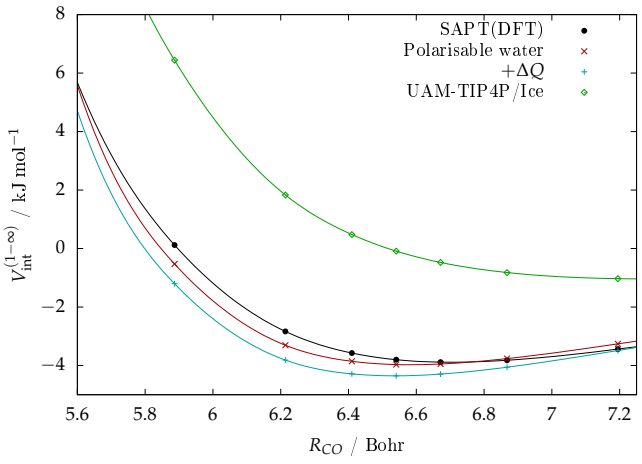
Potentials which combine the UAM potential with a charged-atom water model using LJ mixing rules are common practice, Das et al. used the SPC/E water model alongside LJ "united atom" guests representing a range of molecules from hydrogen to propane using this approach [95]. It should be noted that the methane-water interactions used for nucleation simulation developed by Jacobson and Molinero [89] are intended to work consistently with their own water and methane models in clathrates; and in particular to give the correct solubility of methane in water; but still have weaker interactions than their methane-methane terms (0.63 to 1.26 kJ mol$^{-1}$ vs. 1.42 kJ mol$^{-1}$). Docherty et al. altered the methane-water term by 7%, again leaving it slightly weaker than the methane-methane potential. As indicated in Fig 5.17, the difference might not be too significant: If typical methane-methane separations are close to

8 Bohr then they are configurations for which the methane-water interaction is weaker than methane-methane.

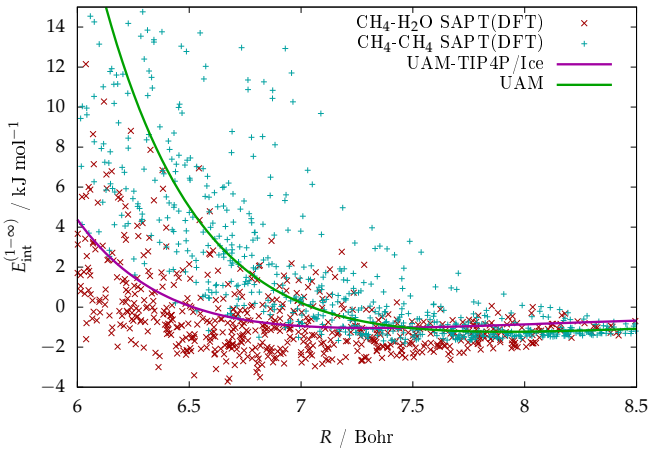
In short, neither of the water-methane potentials used here are without obvious flaws relating to polarisation; our own model probably restricts bubble formation too much whereas the common approach probably lets bubbles form too easily; and rescaling either model might go some way towards fixing the problem.




(a)




(b)




(c)

Figure 5.17: (*a*) Scatter plot and (*b*) profile plot of the global minimum for the water-methane dimer. The plots show the overbinding from replacing the polarisable water model (to which the short-range potential was fitted) with the increased charges. The UAM-TIP4P/Ice interaction from Lorentz-Berthelot combining rules is also shown, this has a depth of about 1.1 kJ mol$^{-1}$ and is too weak an interaction. (*c*) Reference energies plotted against distance for both the CH4 and CH$_4$-H$_2$O dimers, showing that the water-methane and methane-methane interactions tend to be quite close for separations beyond about 7.5 Bohr, but below this separation the methane-water interaction tends to be lower.

# Chapter 6

# Summary

The long-term goal of this thesis was to produce a set of potentials which could be used for simulations of methane clathrate, and in particular to be able to answer current questions about the nucleation and decomposition process. The first step towards this was to produce as elaborate a model as possible for water and methane, and water-methane interactions. Although not currently usable in molecular dynamics simulations, such potentials may be usable in the near-future. The second step was to take these potentials, and use them as a starting point for far simpler potentials which could run without issue in DL_POLY4. These were used in separate simulations of water and methane, and finally in simulations of clathrate decomposition.

## 6.1 Starting Models

There are already a number of sophisticated water models; ours differed from others in the level of the polarisation model, the inclusion of multipoles made using the recently developed DF-ISA method, and the inclusion of shape functions up to terms including the $C_{22c}$ spherical harmonic.

It was found during the fitting process that the multipoles made using the DMA2 distributed multipole method did not behave correctly. In particular for the water-methane interaction the electrostatic energy calculated using DMA multipoles did not converge with the reference electrostatic energy from SAPT(DFT), even at large separations it was only around half the reference energy. The multipole model chosen gave about $1.1E^{(1)}_{\text{elst}}$ for methane-water, and $0.98E^{(1)}_{\text{elst}}$ for water-water interactions at large separations.

Early attempts at fitting the water pair potential failed to find the minimum, and this was because the shape function lacked a $\rho_{22c}$ term on the oxygen site, which allows deformation of the repulsive wall in the out-of-plane direction of the molecule. When this term was allowed in the fit it was assigned a prefactor an order of magnitude greater than the existing shape function terms (the $\rho_{11c}$ term was also allowed on the hydrogen site but remained small and so probably had little effect). It might be taken that the improvements with the $\rho_{22c}$ term come from accounting for the lone-pair electrons of the oxygen atom.

With the new multipoles and shape function terms included; the model fitted the reference energies so readily that the fitting method could be simplified: An intermediate step where the first-order energy was fitted to the overlap model could now be removed from the process.

The final model gave a dimer minimum as good or better than data from other available models. It was also used to give reminimised energies of water clusters, for which it performed comparably well to existing polarisable models. It may be worth noting that the potential was fitted to around 2000 SAPT(DFT) reference energies, whereas there are no SAPT(DFT) reference energies for water clusters. Recent water models tend to use a very large number of CCSD(T) reference energies.

Comparatively little insight can be taken from developing the methane model, since methane potentials have been of little interest historically. It can be noted that the dispersion model was not the intended L3 dispersion model calculated from WSM frequency-dependent polarisabilities using CamCASP since this produced a systematic error of around +0.16 kJ $mol^{-1}$, so instead a $C_6$-only potential was fitted to reference energies, which it did quite well. The methane potential fitted the global minimum well except that the separation was too high by around 0.2 Bohr. This would be a consistent problem in the methane modelling.

It should be stressed that the final fitting procedure is really quite straight forward, with minimum demand of qualitative human input: Once the functional form and the multipole model had been decided on, it was only really a single anchor strength and a weighting parameter for the reference energies that had to be chosen by hand, once at first-order and once at infinite-order, to generate the water model. With regards to the anchor strength, it is of course possible in principle to adjust anchor strengths for each term in the short range potential individually and it can only be assumed that a better fit still could be obtained this way. As it is the potentials demonstrate the validity of a method asking for relatively low computational power and few reference points in comparison with other similar water models. The general method used here has since been used with few changes in developing a potential for the pyradine dimer [98].

## 6.2 Models for DL_POLY

Simplifying the methane potential was in a way very straight forward: The shape function was replaced with an isotropic exponential repulsive wall, the multipoles were replaced with the ISA charges, and it was found there was no clear benefit to making it polarisable. An alternative model was made using Lennard-Jones interactions between atoms, this did a much poorer job of matching reference energies or the global minimum although it did give a similar quality of result when used to reminimise small methane clusters. The main problem associated with the simplification was the difficulty in fitting to the global minimum. The final fit was out by around 0.3 Bohr. It was argued that the global minimum profile for the methane dimer is particularly unlike other dimers in that the separation at the minimum is very low, and that in retrospect getting the methane minimum right is probably not that important. Additional sources of error are the charge model, which underwent little consideration and could probably be improved quite easily; and the wide and shallow form of the methane potential, which obviously cannot be helped.

Fitting the water model was more difficult. The multipole model had to be accounted for somehow by the inclusion of extra charges. Early on it was assumed that the extra charges should lie off the oxygen site a way that would intuitively lend itself to the tetrahedral bonding water tends to prefer. Instead, the tetrahedral charge model failed at exactly the angle of

hydrogen bonding. The charge sites of the existing SAPT-5s charge model were used instead, with new charge values fitted. This is not an ideal choice; the SAPT-5s potential was intended to work alongside extra exchange-repulsion sites; but it the arrangement seems to have been sufficient. Since the final model is non-polarisable, there should be no problem making a model using the earlier 7-site charge arrangement or something similar. At first it was assumed that a model with three undamped polarisable sites could be transferred into DL_POLY, but this was not the case. An alternative model with a single polarisable site was tried instead, this model failed to reproduce the energetic ordering of hexamers, and in simulations of ice Ih gave a melting temperature below 120 K. Instead, polarisation was accounted for by increasing the charge values to match a water molecule in dielectric. For MD simulations, this seems like the best method of accounting for polarisation, unless adequate damped polarisation models can be incorporated.

The resulting water model reproduces the ordering of the hexamers well, unlike the empirical TIP4P models, but also reproduces the radial distribution functions for liquid water and for ice. Like many water models it does have an obvious shortcoming underestimating the melting temperature by about 30 to 40 K.

## 6.3 Clathrate decomposition

Simulations on methane clathrate decomposition differed greatly between using our own models and using the common combination of TIP4P/Ice and UAM methane. The discrepancy was apparently due to the water-methane and methane-methane potentials and not the water models. Using UAM and the associated water potential leads to nanobubble formation as the clathrate melts; although the slower the clathrate is melted, the more it is able to melt before bubbles form. This might be seen to support the clathrate memory effect through supersaturation of methane. Under further scrutiny, it was concluded that the conventional methane-water potential in this case is probably too weak, encouraging bubbles to form faster than they would in nature. Conversely, the water-methane potentials developed here were slightly overbinding, as a result of the increased charges on the water model. The result was that when dissolved in water, methane had very little inclination to gather into bubbles at all. In the future, it might be possible to refit the methane-water potential to account for this, this should lead to a more realistic clathrate decomposition picture.

# Appendix A

# Transferring Potentials into DL_POLY4

The potentials to be put into DL_POLY were isotropic but were not of an analytical form available in DL_POLY. In order to replicate them, they had to be included in tabular form. For each pair potential a grid of 10000 points was used, spaced 0.001 Å apart to cover 10 Å total. Entries for the table were made using Mathematica 10.

The electrostatic energy is not included in the table; this comes from the site charges and is handled using Ewald summation. Due to rounding errors the charge models for water that had been produced by Mulfit do not necessarily sum to zero total charge, in this case the charge of the oxygen site was changed. For example, the $+\Delta Q$ charge model has a total charge of +0.0001 AU, so the oxygen charge was changed from -0.13461 to -0.13462. This introduced only small errors, on the order of $10^{-4}$ or $10^{-5}$ for the water-methane dimer.

To test the potentials had been transferred correctly, a few dimers along the global minimum were placed in a cubic box 100 Å wide. An example is given for the water-methane dimer, in which it was found that energies from Orient calculations to DL_POLY calculations differed by around a percent for lower-energy values, most of which was accounted for by the electrostatic energy.

| R / Bohr | $E_{\text{int}}^{(1-\infty)}$ from Orient | $E_{\text{int}}^{(1-\infty)}$ from DL_POLY | $E_{\text{elst}}^{(1)}$ from Orient | $E_{\text{elst}}^{(1)}$ from DL_POLY | Total error | Error from $E_{\text{elst}}^{(1)}$ |
|---|---|---|---|---|---|---|
| 5.23272 | 24.268 | 24.253 | $-9.6597$ | $-9.6743$ | $-0.015$ | $-0.0146$ |
| 5.55977 | 6.1922 | 6.1929 | $-6.7562$ | $-6.7653$ | $+0.007$ | $+0.0091$ |
| 6.54091 | $-4.3537$ | $-4.3560$ | $-2.5908$ | $-2.5932$ | $-0.0013$ | $-0.0024$ |
| 7.52204 | $-2.8834$ | $-2.8844$ | $-1.1431$ | $-1.1441$ | $-0.0010$ | $-0.0010$ |

Table A.1: Total and electrostatic energies in kJ mol$^{-1}$ for points along the profile of the minimum for the water-methane dimer, using Orient and DL_POLY4.

# Appendix B

# Full model specifications

The final sets of anisotropic and isotropic models are given here. The sections relevant to each model are listed in Table B.

| Model | Geometry [axes] | Electrostatics | Polarisation [damping] | Dispersion [damping] | Exchange-Repulsion |
|---|---|---|---|---|---|
| Anisotropic Water | B.1 [B.2.1] | B.4 | B.7 B.8 [B.13] | B.12 [B.13] | B.14 |
| Isotropic (Polarisable) Water | B.1 [B.2.1] | B.6 | (B.9 [B.13]) | B.12 [B.13] | B.17 |
| Anisotropic Methane | B.2 [B.2.2] | B.5 | B.10 B.11 [B.13] | B.12 [B.13] | B.15 |
| Isotropic Methane | B.2 [B.2.2] | B.6 | | B.12 [B.13] | B.17 |
| Anisotropic Water-Methane | B.1,B.2 [B.2.1,B.2.2] | B.4,B.5 | B.7 B.8,B.10 B.11 [B.13] | B.12 [B.13] | B.16 |
| Isotropic (Polarisable) Water-Methane | B.1,B.2 [B.2.1,B.2.2] | B.6 | (B.9 [B.13]) | B.12 [B.13] | B.17 |

## B.1 Monomer Geometries

Coordinates are included for the methane molecule and for the water molecule from that molecule's frame, including the extra sites "L" used in the isotropic model.

## B.2 Global and Local Axes

### B.2.1 Axes for the water model

The multipole model uses global axes. These are indicated by the coordinates given in Table B.1. The polarisabilities use a local axis system centred on each atom. For oxygen, the x-axis points from H1 to H2, the z-axis points from the oxygen site towards the midpoint of H1 and H2. For the hydrogen atoms, the z-axis points along the line from oxygen to that hydrogen, and the x-axis is perpendicular to this pointing loosely towards the other hydrogen. In all cases the y-axis runs into or out from the plane of the molecule to form right-handed set of axes. These are illustrated in Fig B.1.

### B.2.2 Axes for the methane model

The multipole model uses global axes. These are indicated by the coordinates given in Table B.2. The polarisabilities, when used, use a local axis system centred on each atom. For carbon, the z-axis runs from C towards H1, and the x-axis runs parallel to the line from H3 to H2. For each hydrogen atom, the z-axis runs from C towards that atom and the x-axis runs parallel to the line between two other hydrogen atoms given in Table B.3. An example of this is given in Fig B.2.

## B.3 Full Multipoles

Table B.1: Coordinates for sites in the isotropic water model from the molecule's frame, given in Bohr. The anisotropic model used the same site positions but had atomic sites only.

| Site | x | y | z |
|---|---|---|---|
| O | 0.00000000 | 0.00000000 | 0.00000000 |
| H1 | -1.45365196 | 0.00000000 | -1.12168732 |
| H2 | 1.45365196 | 0.00000000 | -1.12168732 |
| L1 | 0.00000000 | 0.20672100 | -0.37179200 |
| L2 | 0.00000000 | -0.20672100 | -0.37179200 |

Table B.2: Coordinates for sites in all methane models from the molecule's frame, given in Bohr.

| Site | x | y | z |
|---|---|---|---|
| C | 0.00000000 | 0.00000000 | 0.00000000 |
| H1 | 0.00000000 | 2.07703578 | 0.00000000 |
| H2 | 1.69589261 | -0.69234525 | 0.97912406 |
| H3 | -1.69589261 | -0.69234525 | 0.97912406 |
| H4 | 0.00000000 | -0.69234525 | -1.95824810 |

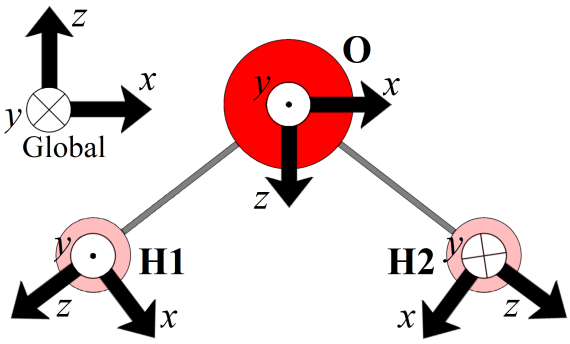



Figure B.1: The global axes and local axes for the water molecule.

Table B.3: The direction of the x-axis for each local axis system. For all hydrogen atoms the z-axis points along the line from C to that atom. For carbon, the z-axis points from C to H1. The y-axis is perpendicular to these to form a right-handed system.

| Atom | x-axis direction |
|---|---|
| C | H3 to H2 |
| H1 | H3 to H2 |
| H2 | H4 to H1 |
| H3 | H1 to H4 |
| H4 | H2 to H3 |

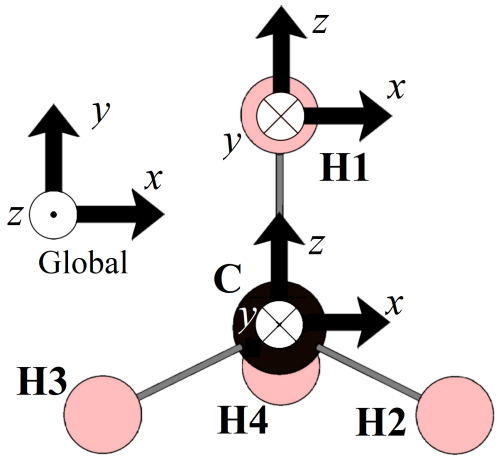



Figure B.2: The global axes and an example of the local axes for the methane molecule.

Table B.4: Non-zero components of the DF-ISA rank 4 multipole model for water.

| Oxygen | |
|---|---|
| $t$ | $Q_t$ |
| 00 | −0.802223 |
| 10 | 0.170838 |
| 20 | −0.021976 |
| 22c | 0.488908 |
| 30 | −0.094213 |
| 32c | 0.130864 |
| 40 | −0.378289 |
| 42c | 0.442066 |
| 44c | 0.085999 |

| Hydrogen | |
|---|---|
| $t$ | $Q_t$ |
| 00 | 0.401111 |
| 10 | −0.005473 |
| 11c | −0.030195 |
| 20 | 0.005473 |
| 21c | 0.019693 |
| 22c | 0.021928 |
| 30 | −0.012700 |
| 31c | −0.021306 |
| 32c | −0.018826 |
| 33c | −0.047698 |
| 40 | 0.050446 |
| 41c | −0.007140 |
| 42c | −0.076937 |
| 43c | −0.038622 |
| 44c | −0.086243 |

Table B.5: Non-zero components of the DF-ISA rank 4 multipole model for methane.

| Carbon | |
|---|---|
| $t$ | $Q_t$ |
| 00 | −0.413530 |
| 10 | −0.000094 |
| 11s | −0.000333 |
| 20 | −0.000072 |
| 21s | 0.000048 |
| 22c | 0.000132 |
| 30 | 0.129917 |
| 31s | 0.111463 |
| 32c | −0.099948 |
| 33s | 0.143096 |
| 40 | 0.659155 |
| 41s | 1.967992 |
| 42c | 0.986644 |
| 43s | −2.231196 |
| 44c | 1.302589 |
| Hydrogen | |
| $t$ | $Q_t$ |
| 00 | 0.103598 |
| 10 | −0.000019 |
| 11s | 0.035237 |
| 20 | −0.001059 |
| 21s | 0.000029 |
| 22c | −0.001804 |
| 30 | −0.114429 |
| 31s | −0.109851 |
| 32c | 0.088540 |
| 33s | −0.142021 |

## B.4 Charge models

Table B.6: Charge models for the final water and methane potentials.

| Water using SAPT-5s site locations | $Q$ | $Q + \Delta Q/2$ | $Q + \Delta Q$ |
|---|---|---|---|
| O | 0.11923 | -0.00882 | -0.13461 |
| H | 0.52245 | 0.56041 | 0.59852 |
| L | -0.59207 | -0.55599 | -0.53121 |
| Methane | | | |
| C | -0.413530 | | |
| H | 0.103598 | | |

## B.5 Polarisabilties

### B.5.1 Water Polarisabilities

Table B.7: Components of the polarisability tensor $\alpha_{l,m}^{O,O}$ for the water molecule used in the anisotropic models, given in atomic units.

| l,m | $\alpha_{l,m}$ | l,m | $\alpha_{l,m}$ |
|---|---|---|---|
| 10,10 | 5.8757934 | 21c,33c | -7.1092025 |
| 10,20 | -3.3897288 | 21s,21s | 33.301433 |
| 10,22c | 2.8974678 | 21s,31s | 10.990663 |
| 10,30 | -3.6365435 | 21s,33s | -19.790110 |
| 10,32c | -8.6221441 | 22c,22c | 24.904919 |
| 11c,11c | 6.5000178 | 22c,30 | 20.262685 |
| 11c,21c | 3.2732374 | 22c,31c | -1.0083506 |
| 11c,31c | 13.822410 | 22c,32c | 0.41659226 |
| 11c,33c | -14.844736 | 22c,33c | -0.30892309 |
| 11s,11s | 5.6509062 | 22s,22s | 22.772251 |
| 11s,21s | -2.1500681 | 22s,32s | -14.019709 |
| 11s,31s | -4.5122123 | 30,30 | 252.01573 |
| 11s,33s | -17.489275 | 30,31c | 2.7351932 |
| 20,20 | 18.858820 | 30,32c | 49.573411 |
| 20,22c | 15.067780 | 30,33c | 1.4971095 |
| 20,30 | 0.78561129 | 31c,31c | 207.33784 |
| 20,31c | 0.82200040 | 31c,32c | -3.4739037 |
| 20,32c | 1.4633191 | 31c,33c | 38.493651 |
| 20,33c | 0.16764731 | 31s,31s | 285.22531 |
| 21c,21c | 20.911721 | 31s,33s | 6.3765611 |
| 21c,22c | -0.061418394 | 32c,32c | 236.30645 |
| 21c,30 | -0.80603401 | 32s,32s | 279.81398 |
| 21c,31c | -4.1039635 | 33c,33c | 283.47721 |
| 21c,32c | 0.35142101 | 33s,33s | 244.53973 |

Table B.8: Components of the polarisability tensor $\alpha_{l,m}^{H,H}$ for the water molecule used in the anisotropic models, given in atomic units.

| l,m | $\alpha_{l,m}$ |
|---|---|
| 10,10 | 1.7817493 |
| 10,11c | 0.037455169 |
| 11c,11c | 1.8285965 |
| 11s,11s | 1.7477126 |

Table B.9: Local polarisabilities for the water molecule used in the isotropic models, given in atomic units.

| atom | $\alpha$ |
|---|---|
| O | 5.40801522 |
| H | 1.60741755 |

### B.5.2 Methane Polarisabilities

Table B.10: Components of the polarisability tensor $\alpha_{l,m}^{C,C}$ for the methane molecule used in the anisotropic models.

| l,m | $\alpha_{l,m}$ | l,m | $\alpha_{l,m}$ |
|---|---|---|---|
| 10,10 | 5.8065524 | 21c,33c | 0.39175720 |
| 10,20 | 13.751560 | 21s,21s | 28.659115 |
| 10,21s | 0.020723762 | 21s,22c | 14.066607 |
| 10,22c | -0.031357044 | 21s,30 | 0.54044578 |
| 10,30 | 17.116401 | 21s,31s | -8.6504498 |
| 10,31s | -0.016966915 | 21s,32c | 10.040091 |
| 10,32c | -0.031550471 | 21s,33s | -0.31297283 |
| 10,33s | 18.855072 | 22c,22c | 38.404614 |
| 11c,11c | 5.8124463 | 22c,30 | -0.14704728 |
| 11c,21c | -7.9232893 | 22c,31s | -5.8764630 |
| 11c,22s | -11.141826 | 22c,32c | 17.115600 |
| 11c,31c | -10.072684 | 22c,33s | -0.42313078 |
| 11c,32s | 22.900059 | 22s,22s | 38.571126 |
| 11c,33c | -0.21055731 | 22s,31c | -7.6608001 |
| 11s,11s | 5.8087242 | 22s,32s | 13.821085 |
| 11s,21s | -7.9479926 | 22s,33c | -0.81694747 |
| 11s,22c | -11.250549 | 30,30 | 1030.0176 |
| 11s,30 | -0.16131281 | 30,31s | -3.4409967 |
| 11s,31s | -10.437666 | 30,32c | 10.999123 |
| 11s,32c | 22.675566 | 30,33s | 94.227132 |
| 11s,33s | -0.51220132 | 31c,31c | 1018.1065 |
| 20,20 | 48.126365 | 31c,32s | -52.070129 |
| 20,21s | 0.16489452 | 31c,33c | 6.7006551 |
| 20,22c | 0.084121334 | 31s,31s | 986.27409 |
| 20,30 | -14.295966 | 31s,32c | -54.898302 |
| 20,31s | 2.7314899 | 31s,33s | -2.7669889 |
| 20,32c | 0.45825582 | 32c,32c | 1115.0271 |
| 20,33s | -17.765002 | 32c,33s | -10.346532 |
| 21c,21c | 28.828128 | 32s,32s | 1107.9736 |
| 21c,22s | 13.625172 | 32s,33c | 20.421240 |
| 21c,31c | -4.3430588 | 33c,33c | 984.05994 |
| 21c,32s | 11.365481 | 33s,33s | 1026.3990 |

Table B.11: Components of the polarisability tensor $\alpha_{l,m}^{H,H}$ for the methane molecule used in the anisotropic models.

| l,m | $\alpha_{l,m}$ |
|---|---|
| 10,10 | 2.6739317 |
| 11c,11c | 3.0704533 |
| 11s,11s | 3.0780352 |

## B.6 Dispersion Terms

Table B.12: Dispersion coefficients for all pairs, in atomic units.

| Pair | $C_6$ | $C_8$ | $C_{10}$ | $C_{12}$ |
|---|---|---|---|---|
| Water | | | | |
| O,O | 18.65100 | 404.1823 | 11886.55 | 253436.4 |
| O,Hw | 5.195981 | 55.85792 | 1081.399 | |
| Hw,Hw | 1.480481 | | | |
| Methane | | | | |
| C,C | 22.07923 | | | |
| C,H | 10.09522 | | | |
| H,H | 4.87260 | | | |
| Water-methane | | | | |
| O,C | 16.90574 | 497.2595 | 24469.86 | 668674.5 |
| O,H | 8.09713 | 85.56545 | 1658.857 | |
| Hw,C | 4.77852 | 87.49391 | 4100.564 | |
| Hw,H | 2.47039 | | | |

## B.7 Dispersion and Polarisation Damping

Table B.13: Dispersion and polarisation Tang-Toennies damping parameters for all pairs.

| Pair | $\beta_{\text{disp}}$ | $\beta_{\text{pol}}$ |
|---|---|---|
| Water | | |
| O,O | 1.89 | 1.08 |
| O,Hw | 1.89 | 1.47 |
| Hw,Hw | 1.89 | 2.00 |
| Methane | | |
| C,C | 1.93 | 1.93 |
| C,H | 1.93 | 1.93 |
| H,H | 1.93 | 1.93 |
| Water-Methane | | |
| O,C | 1.88 | 1.35 |
| O,H | 1.88 | 1.35 |
| Hw,C | 1.88 | 1.35 |
| Hw,H | 1.88 | 1.35 |

## B.8 Short Range Terms

Table B.14: Short-range parameters for the anisotropic water model.

| Pair | $\alpha$ | $l_a$ | $l_b$ | $\rho$ |
|---|---|---|---|---|
| O,O | 1.92296 | 00 | 00 | 5.725084 |
| | | 10 | | 0.000441 |
| | | 20 | | 0.009604 |
| | | 22c | | −0.100775 |
| O,H | 1.878702 | 00 | 00 | 4.775246 |
| | | 10 | | 0.004987 |
| | | | 10 | −0.277651 |
| | | | 11c | 0.033224 |
| | | 20 | | 0.015348 |
| | | 22c | | −0.178707 |
| H,H | 1.932032 | 00 | 00 | 3.677247 |
| | | 10 | | −0.123442 |
| | | 11c | | −0.002921 |

Table B.15: Short-range parameters for the anisotropic methane model.

| Pair | $\alpha$ | $l_a$ | $l_b$ | $\rho$ |
|---|---|---|---|---|
| C,C | 2.179220 | 00 | 00 | 5.552955 |
| C,H | 1.898729 | 00 | 00 | 4.186650 |
| | | | 10 | −0.204299 |
| H,H | 1.647922 | 00 | 00 | 4.782582 |
| | | 10 | | −0.275546 |

Table B.16: Short-range parameters for the anisotropic water-methane model.

| Pair | $\alpha$ | $l_a$ | $l_b$ | $\rho$ |
|---|---|---|---|---|
| O,C | 2.206657 | 00 | 00 | 5.375834 |
| | | 10 | | −0.015507 |
| | | 20 | | −0.064900 |
| | | 22c | | −0.151909 |
| O,H | 1.695928 | 00 | 00 | 5.265648 |
| | | 10 | | −0.002486 |
| | | | 10 | −0.331392 |
| | | 20 | | 0.015588 |
| | | 22c | | −0.144187 |
| Hw,C | 1.688009 | 00 | 00 | 4.769332 |
| | | 10 | | −0.100053 |
| | | 11c | | 0.081363 |
| Hw,H | 1.891707 | 00 | 00 | 4.109502 |
| | | 10 | | −0.292882 |
| | | | 10 | −0.238494 |
| | | 11c | | 0.009255 |

## B.9 Commonly used abbreviations

Table B.17: Short-range parameters for the isotropic models.

| Pair | $\alpha$ | $\rho$ |
|---|---|---|
| Water | | |
| O,O | 1.844988 | 6.012939 |
| O,Hw | 2.269041 | 4.064269 |
| Hw,Hw | 1.987949 | 4.314218 |
| Methane | | |
| C,C | 1.821205 | 6.379491 |
| C,H | 1.695672 | 4.825019 |
| H,H | 1.576686 | 4.302228 |
| Water-methane | | |
| O,C | 2.044257 | 6.338081 |
| O,H | 2.143781 | 4.550381 |
| Hw,C | 2.143781 | 4.026565 |
| Hw,H | 1.710300 | 3.814876 |

Table B.18: List of frequently-used abbreviations

| | |
|---|---|
| SAPT | Symmetry-Adapted Perturbation Theory |
| DFT | Density Functional Theory |
| CCSD(T) | Coupled Cluster with single and double excitations, and triple excitations described through perturbation. |
| MP2 | Møller–Plesset theory with second-order. perturbations |
| CBS | Complete Basis Set |
| ISA | Iterated Stockholder Atom method; used to produce multipoles. |
| RDF | Radial Distribution Function |
| L3 (Dispersion) | A dispersion model with terms up to $C_{12}/R^{12}$. |
| L4 (Multipoles) | A multipole model with terms up to hexadecapole. |
| Aniso[tropic] model | Used to refer to a model developed here with anisotropic exchange-repulsion and anisotropic polarisation. |
| Iso[tropic] model | Used to refer to a model developed here with only isotropic exchange-repulsion and isotropic polarisation. |

# Bibliography


[1] B. C. Barnes, A. K. Sum, *Current Opinion in Chemical Engineering* **2013**.

[2] I. Chatti, A. Delahaye, L. Fournaison, J.-P. Petitet, *Energy Conversion and Management* **2005**, *46*, 1333–1343.

[3] P. Ren, J. W. Ponder, *The Journal of Physical Chemistry B* **2003**, *107*, 5933–5947.

[4] P. Ren, J. W. Ponder, *The Journal of Physical Chemistry B* **2004**, *108*, 13427–13437.

[5] P. Ren, C. Wu, J. W. Ponder, *Journal of chemical theory and computation* **2011**, *7*, 3143–3161.

[6] Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder, P. Ren, *Journal of chemical theory and computation* **2013**, *9*, 4046–4063.

[7] R. Bukowski, K. Szalewicz, G. C. Groenenboom, A. Van der Avoird, *Science* **2007**, *315*, 1249–1252.

[8] S. S. L. Price, L. S. Price in *Intermolecular Forces and Clusters I*, Springer, **2005**, pp. 81–123.

[9] S. S. L. Price, *Accounts of chemical research* **2008**, *42*, 117–126.

[10] M. J. Van Vleet, A. J. Misquitta, A. J. Stone, J. R. Schmidt, *Journal of chemical theory and computation* **2016**, *12*, 3851–3870.

[11] J. Bernal, R. Fowler, *The Journal of Chemical Physics* **1933**, *1*, 515–548.

[12] J. L. Abascal, C. Vega, *The Journal of chemical physics* **2005**, *123*, 234505.

[13] J. Abascal, E Sanz, R Garcia Fernandez, C Vega, *The Journal of chemical physics* **2005**, *122*, 234511.

[14] E. M. Mas, R. Bukowski, K. Szalewicz, G. C. Groenenboom, P. E. Wormer, A. van der Avoird, *The Journal of Chemical Physics* **2000**, *113*, 6687.

[15] E. M. Mas, K. Szalewicz, R. Bukowski, B. Jeziorski, *The Journal of chemical physics* **1997**, *107*, 4207–4218.

[16] W. Cencek, K. Szalewicz, C. Leforestier, R. Van Harrevelt, A. van der Avoird, *Physical Chemistry Chemical Physics* **2008**, *10*, 4716–4731.

[17] K. Szalewicz, G. Murdachaew, R. Bukowski, O. Akin-Ojo, C. Leforestier in Lecture Series on Computer and Computational Science: International Conference on Computational Methods in Science and Engineering (ICCMSE 2006), *Vol. 6*, **2006**, pp. 482–491.

[18] C. J. Burnham, J. Li, S. S. Xantheas, M. Leslie, *The Journal of chemical physics* **1999**, *110*, 4566–4581.

[19] C. J. Burnham, S. S. Xantheas, *The Journal of chemical physics* **2002**, *116*, 1479–1492.

[20] S. S. Xantheas, C. J. Burnham, R. J. Harrison, *The Journal of chemical physics* **2002**, *116*, 1493–1499.

[21] C. J. Burnham, S. S. Xantheas, *The Journal of chemical physics* **2002**, *116*, 1500–1510.

[22] C. J. Burnham, S. S. Xantheas, *The Journal of chemical physics* **2002**, *116*, 5115–5124.

[23] G. S. Fanourgakis, S. S. Xantheas, *The Journal of Physical Chemistry A* **2006**, *110*, 4100–4106.

[24] G. S. Fanourgakis, S. S. Xantheas, *The Journal of chemical physics* **2008**, *128*, 074506.

[25] C. Burnham, D. Anick, P. Mankoo, G. Reiter, *The Journal of chemical physics* **2008**, *128*, 154519.

[26] V. Babin, G. R. Medders, F. Paesani, *Journal of chemical theory and computation* **2014**, *10*, 1599–1607.

[27] K. T. Wikfeldt, E. Batista, F. Vila, H Jónsson, *Physical Chemistry Chemical Physics* **2013**, *15*, 16542–16556.

[28] M. E. Johnson, T. Head-Gordon, A. A. Louis, *The Journal of chemical physics* **2007**, *126*, 144509.

[29] H. Wang, C. Junghans, K. Kremer, *The European Physical Journal E: Soft Matter and Biological Physics* **2009**, *28*, 221–229.

[30] V. Molinero, E. B. Moore, *The Journal of Physical Chemistry B* **2008**, *113*, 4008–4016.

[31] S. Riniker, W. F. van Gunsteren, *The Journal of chemical physics* **2011**, *134*, 084110.

[32] S. O. Yesylevskyy, L. V. Schäfer, D. Sengupta, S. J. Marrink, *PLoS computational biology* **2010**, *6*, e1000810.

[33] W. L. Jorgensen, J. D. Madura, C. J. Swenson, *Journal of the American Chemical Society* **1984**, *106*, 6638–6646.

[34] L. Verlet, J.-J. Weis, *Molecular Physics* **1972**, *24*, 1013–1024.

[35] S. J. Goodbody, K. Watanabe, D. MacGowan, J. P. Walton, N. Quirke, *Journal of the Chemical Society Faraday Transactions* **1991**, *87*, 1951–1958.

[36] B Guillot, Y Guissani, *The Journal of chemical physics* **1993**, *99*, 8075–8094.

[37] J. S. Tse, M. L. Klein, I. R. McDonald, *The Journal of Physical Chemistry* **1983**, *87*, 4198–4203.

[38] S Murad, K. Gubbins in, ACS Publications, **1978**.

[39] Z. Cao, J. W. Tester, K. A. Sparks, B. L. Trout, *The Journal of Physical Chemistry B* **2001**, *105*, 10950–10960.

[40] O. Akin-Ojo, K. Szalewicz, *The Journal of chemical physics* **2005**, *123*, 134311.

[41] C. Qu, R. Conte, P. L. Houston, J. M. Bowman, *Physical Chemistry Chemical Physics* **2015**, *17*, 8172–8181.

[42] C. McKay, K. Hand, P. Doran, D. Andersen, J. Priscu, *Geophysical Research Letters* **2003**, *30*, 1702.

[43] B. Buffett, D. Archer, *Earth and Planetary Science Letters* **2004**, *227*, 185–199.

[44] Active seafloor gas vents on the shelf and upper slope in Canadian Beaufort Sea, American Geophysical Union Fall Meeting, **2012**.

[45] E. D. Sloan, *Energy & Fuels* **1998**, *12*, 191–196.

[46] J. Mienert, M. Vanneste, S. Bünz, K. Andreassen, H. Haflidason, H. P. Sejrup, *Marine and Petroleum Geology* **2005**, *22*, 233–244.

[47] U. Marboeuf, O. Mousis, J.-M. Petit, B. Schmitt, *The Astrophysical Journal* **2010**, *708*, 812.

[48] O. Mousis, J. I. Lunine, K. E. Mandt, E. Schindhelm, H. A. Weaver, S. A. Stern, J. H. Waite, R. Gladstone, A. Moudens, *Icarus* **2013**.

[49] S. L. Miller, W. D. Smythe, *Science* **1970**, *170*, 531–533.

[50] A Dobrovolskis, A. Ingersoll, *Icarus* **1975**, *26*, 353–357.

[51] D. Musselwhite, J. I. Lunine, *Journal of Geophysical Research: Planets (1991–2012)* **1995**, *100*, 23301–23306.

[52] E. Chassefière, E. Dartois, J.-M. Herri, F. Tian, F. Schmidt, O. Mousis, A. Lakhlifi, *Icarus* **2013**.

[53] D Toublanc, J. Parisot, J Brillet, D Gautier, F Raulin, C. McKay, *Icarus* **1995**, *113*, 2–26.

[54] M. Choukroun, O. Grasset, G. Tobie, C. Sotin, *Icarus* **2010**, *205*, 581–593.

[55] A. Falenty, T. C. Hansen, W. F. Kuhs, *Nature* **2014**, *516*, 231–233.

[56] L. Del Rosso, M. Celli, L. Ulivi, *Nature communications* **2016**, *7*, 13394.

[57] D. Davidson, S. Gough, J. Ripmeester, H. Nakayama, *Canadian Journal of Chemistry* **1981**, *59*, 2587–2590.

[58] P. M. Rodger, *Journal of physical chemistry* **1990**, *94*, 6080–6089.

[59] P. Englezos, *Industrial & Engineering Chemistry Research* **1993**, *32*, 1251–1274.

[60] C. Moon, P. C. Taylor, P. M. Rodger, *Journal of the American Chemical Society* **2003**, *125*, 4706–4707.

[61] J. A. Ripmeester, S. T. John, C. I. Ratcliffe, B. M. Powell, *Nature* **1987**, *325*, 135–136.

[62] K. A. Udachin, C. I. Ratcliffe, J. A. Ripmeester, *Angewandte Chemie* **2001**, *113*, 1343–1345.

[63] R. McMullan, G. Jeffrey, T. H. Jordan, *The Journal of Chemical Physics* **1967**, *47*, 1229.

[64] Japan extracts gas from methane hydrate in world first, **2013**.

[65] China claims breakthrough in mining flammable ice, **2013**.

[66] J. F. Gabitto, C. Tsouris, *Journal of Thermodynamics* **2010**, *2010*.

[67] L Ulivi, M Celli, A Giannasi, A. Ramirez-Cuesta, D. Bull, M Zoppi, *Physical Review B* **2007**, *76*, 161401.

[68] C. K. Paull, W. Ussler III, S. R. Dallimore, S. M. Blasco, T. D. Lorenson, H. Melling, B. E. Medioli, F. M. Nixon, F. A. McLaughlin, *Geophysical Research Letters* **2007**, *34*, L01603.

[69] L. C. Jacobson, W. Hujo, V. Molinero, *The Journal of Physical Chemistry B* **2009**, *113*, 10298–10307.

[70] S. Alavi, J. Ripmeester, *Journal of Chemical Physics* **2010**, *132*, 144703–1.

[71] T. Yagasaki, M. Matsumoto, Y. Andoh, S. Okazaki, H. Tanaka, *The Journal of Physical Chemistry B* **2014**, *118*, 1900–1906.

[72] S. A. Bagherzadeh, S. Alavi, J. A. Ripmeester, P. Englezos, *Fluid Phase Equilibria* **2013**, *358*, 114–120.

[73] R. L. Christiansen, E. D. Sloan, *Annals of the New York Academy of Sciences* **1994**, *715*, 283–305.

[74] R. Radhakrishnan, B. L. Trout, *The Journal of chemical physics* **2002**, *117*, 1786–1796.

[75] G. Jeffrey, J. Atwood, J. Davies, D. MacNicol, *Edited by* **1984**, 135.

[76] J. A. Ripmeester, S. Alavi, *Current Opinion in Solid State and Materials Science* **2016**.

[77] P. Wilson, A. Haymet, *Chemical engineering journal* **2010**, *161*, 146–150.

[78] H. Jiang, K. D. Jordan, C. Taylor, *The Journal of Physical Chemistry B* **2007**, *111*, 6486–6492.

[79] H. Jiang, E. M. Myshakin, K. D. Jordan, R. P. Warzinski, *The Journal of Physical Chemistry B* **2008**, *112*, 10207–10216.

[80] H. Yu, W. F. van Gunsteren, *The Journal of chemical physics* **2004**, *121*, 9549.

[81] L. Xu, X. Wang, L. Liu, M. Yang, *Computational and Theoretical Chemistry* **2011**, *977*, 209–212.

[82] R. W. Hawtin, D. Quigley, P. M. Rodger, *Physical Chemistry Chemical Physics* **2008**, *10*, 4853–4864.

[83] M. R. Walsh, C. A. Koh, E. D. Sloan, A. K. Sum, D. T. Wu, *Science* **2009**, *326*, 1095–1098.

[84] L. C. Jacobson, W. Hujo, V. Molinero, *The Journal of Physical Chemistry B* **2010**, *114*, 13796–13807.

[85] L. C. Jacobson, W. Hujo, V. Molinero, *Journal of the American Chemical Society* **2010**, *132*, 11806–11811.

[86] T. Koga, J. Wong, M. K. Endoh, D. Mahajan, C. Gutt, S. K. Satija, *Langmuir* **2010**, *26*, 4627–4630.

[87] M. R. Walsh, J. D. Rainey, P. G. Lafond, D.-H. Park, G. T. Beckham, M. D. Jones, K.-H. Lee, C. A. Koh, E. D. Sloan, D. T. Wu, et al., *Physical Chemistry Chemical Physics* **2011**, *13*, 19951–19959.

[88] L. Tang, Y. Su, Y. Liu, J. Zhao, R. Qiu, *The Journal of Chemical Physics* **2012**, *136*, 224508.

[89] L. C. Jacobson, V. Molinero, *The Journal of Physical Chemistry B* **2010**, *114*, 7302–7311.

[90] B. Beeskow-Strauch, J. M. Schicks, E. Spangenberg, J. Erzinger, *Chemistry-A European Journal* **2011**, *17*, 4376–4384.

[91] M. Zi, D. Chen, H. Ji, G. Wu, *Energy & Fuels* **2016**, *30*, 5643–5650.

[92] P. Pirzadeh, P. G. Kusalik, *Journal of the American Chemical Society* **2013**.

[93] L. C. Jacobson, M. Matsumoto, V. Molinero, *The Journal of chemical physics* **2011**, *135*, 074501.

[94] G.-J. Guo, Y.-G. Zhang, C.-J. Liu, K.-H. Li, *Physical Chemistry Chemical Physics* **2011**, *13*, 12048–12057.

[95] S. Das, V. S. Baghel, S. Roy, R. Kumar, *Physical Chemistry Chemical Physics* **2015**, *17*, 9509–9518.

[96] S. A. Bagherzadeh, S. Alavi, J. Ripmeester, P. Englezos, *The Journal of chemical physics* **2015**, *142*, 214701.

[97] S. A. Bagherzadeh, P. Englezos, S. Alavi, J. A. Ripmeester, *The Journal of Chemical Thermodynamics* **2012**, *44*, 13–19.

[98] A. J. Misquitta, A. J. Stone, *Journal of chemical theory and computation* **2016**, *12*, 4184–4208.

[99] CamCASP: a program for studying intermolecular interactions and for the calculation of molecular properties in distributed form, **2017**.

[100] Orient, **2017**.

[101] A. J. Misquitta, K. Szalewicz, *Chemical physics letters* **2002**, *357*, 301–306.

[102] A. J. Misquitta, B. Jeziorski, K. Szalewicz, *Physical review letters* **2003**, *91*, 033201.

[103] A. J. Misquitta, K. Szalewicz, *The Journal of chemical physics* **2005**, *122*, 214109.

[104] A. J. Misquitta, R. Podeszwa, B. Jeziorski, K. Szalewicz, *The Journal of chemical physics* **2005**, *123*, 214103.

[105] R. Podeszwa, W. Cencek, K. Szalewicz, *Journal of chemical theory and computation* **2012**, *8*, 1963–1969.

[106] D. J. Tozer, N. C. Handy, *The Journal of chemical physics* **1998**, *109*, 10180–10189.

[107] A. J. Misquitta, A. J. Stone, *Journal of Chemical Theory and Computation* **2008**, *4*, 7–18.

[108] A. J. Misquitta, *Journal of Chemical Theory and Computation* **2013**, *9*, 5313–5326.

[109] A. Stone, *The theory of intermolecular forces*, Oxford University Press, **2013**.

[110] K. Tang, J. P. Toennies, *The Journal of chemical physics* **1984**, *80*, 3726–3741.

[111] A. J. Stone, *Journal of Chemical Theory and Computation* **2005**, *1*, 1128–1132.

[112] A. J. Misquitta, A. J. Stone, F. Fazeli, *Journal of Chemical Theory and Computation* **2014**.

[113] T. C. Lillestolen, R. J. Wheatley, *Chemical Communications* **2008**, 5909–5911.

[114] A. J. Misquitta, A. J. Stone, S. L. Price, *Journal of Chemical Theory and Computation* **2008**, *4*, 19–32.

[115] A. J. Misquitta, A. J. Stone, *Molecular Physics* **2008**, *106*, 1631–1643.

[116] M. P. Hodges, R. J. Wheatley, *Chemical Physics Letters* **2000**, *326*, 263–268.

[117] D. J. Wales, J. P. Doye, *The Journal of Physical Chemistry A* **1997**, *101*, 5111–5116.

[118] R. Kumar, F.-F. Wang, G. R. Jenness, K. D. Jordan, *The Journal of chemical physics* **2010**, *132*, 014309.

[119] D. M. Bates, G. S. Tschumper, *The Journal of Physical Chemistry A* **2009**, *113*, 3555–3559.

[120] G. Ferenczy, C. Reynolds, P. Winn, A. Stone, *May be obtained by contacting AJ Stone email address: ajs1@ cam. ac. uk*.

[121] N. De Leeuw, S. Parker, *Physical Review B* **1998**, *58*, 13901.

[122] P. J. Van Maaren, D. Van Der Spoel, *The Journal of Physical Chemistry B* **2001**, *105*, 2618–2626.

[123] G. Lamoureux, A. D. MacKerell Jr, B. Roux, *The Journal of chemical physics* **2003**, *119*, 5185–5197.

[124] G. Lamoureux, E. Harder, I. V. Vorobyov, B. Roux, A. D. MacKerell, *Chemical Physics Letters* **2006**, *418*, 245–249.

[125] W. Yu, P. E. Lopes, B. Roux, A. D. MacKerell Jr, *The Journal of chemical physics* **2013**, *138*, 034508.

[126] D. P. Fernández, Y Mulev, A. Goodwin, J. L. Sengers, *Journal of Physical and Chemical Reference Data* **1995**, *24*, 33–70.

[127] H. Berendsen, J. Grigera, T. Straatsma, *Journal of Physical Chemistry* **1987**, *91*, 6269–6271.

[128] G. W. Welch, P. G. Karamertzanis, A. J. Misquitta, A. J. Stone, S. L. Price, *Journal of chemical theory and computation* **2008**, *4*, 522–532.

[129] M. W. Mahoney, W. L. Jorgensen, *The Journal of Chemical Physics* **2000**, *112*, 8910–8922.

[130] R. Bukowski, K. Szalewicz, G. C. Groenenboom, A. van der Avoird, *The Journal of chemical physics* **2008**, *128*, 094313.

[131] G. S. Tschumper, M. L. Leininger, B. C. Hoffman, E. F. Valeev, H. F. Schaefer III, M. Quack, *The Journal of chemical physics* **2002**, *116*, 690–701.

[132] A. H.-T. Li, S. D. Chao, *The Journal of chemical physics* **2006**, *125*, 094312.

[133] S. Yoo, S. S. Xantheas in *Handbook of Computational Chemistry*, Springer, **2012**, pp. 761–792.

[134] J. A. Anderson, K. Crager, L. Fedoroff, G. S. Tschumper, *The Journal of chemical physics* **2004**, *121*, 11023–11029.

[135] C. Pérez, M. T. Muckle, D. P. Zaleski, N. A. Seifert, B. Temelso, G. C. Shields, Z. Kisiel, B. H. Pate, *Science* **2012**, *336*, 897–901.

[136] M. Losada, S. Leutwyler, *Journal of Chemical Physics* **2002**, *117*, 2003–2016.

[137] G. A. Cisneros, K. T. Wikfeldt, L. Ojamae, J. Lu, Y. Xu, H. Torabifard, A. P. Bartok, G. Csanyi, V. Molinero, F. Paesani, *Chemical reviews* **2016**, *116*, 7501–7528.

[138] G. R. Medders, A. W. Götz, M. A. Morales, P. Bajaj, F. Paesani, *The Journal of chemical physics* **2015**, *143*, 104102.

[139] U. Hohm, K. Kerl, *Berichte der Bunsengesellschaft für physikalische Chemie* **1991**, *95*, 36–42.

[140] J. Trusler, M Zarari, *The Journal of Chemical Thermodynamics* **1992**, *24*, 973–991.

[141] L. Viererblová, J. Kolafa, *Physical Chemistry Chemical Physics* **2011**, *13*, 19925–19935.

[142] R. G. Fernández, J. L. Abascal, C. Vega, *The Journal of chemical physics* **2006**, *124*, 144506.

[143] J. L. Abascal, R. G. Fernández, C. Vega, M. A. Carignano, *The Journal of chemical physics* **2006**, *125*, 166101.

[144] I.-C. Yeh, G. Hummer, *The Journal of Physical Chemistry B* **2004**, *108*, 15873–15879.

[145] S. Tazi, A. Boţan, M. Salanne, V. Marry, P. Turq, B. Rotenberg, *Journal of Physics: Condensed Matter* **2012**, *24*, 284117.

[146] G. R. Medders, V. Babin, F. Paesani, *Journal of chemical theory and computation* **2014**, *10*, 2906–2910.

[147] M Helbaek, B Hafskjold, D. Dysthe, G. Sørland, *Journal of Chemical & Engineering Data* **1996**, *41*, 598–603.

[148] A. Soper, C. Benmore, *Physical review letters* **2008**, *101*, 065502.

[149] L. B. Skinner, C. Huang, D. Schlesinger, L. G. Pettersson, A. Nilsson, C. J. Benmore, *The Journal of chemical physics* **2013**, *138*, 074506.

[150] G. J. Martyna, M. E. Tuckerman, D. J. Tobias, M. L. Klein, *Molecular Physics* **1996**, *87*, 1117–1157.

[151] P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, et al., *PLoS computational biology* **2017**, *13*, e1005659.

[152] E. D. Sloan Jr, C. Koh, *Clathrate hydrates of natural gases*, CRC press, pp. 147–149.

[153] B. Sowa, N. Maeda, *The Journal of Physical Chemistry A* **2015**, *119*, 10784–10790.

[154] H. Henley, E. Thomas, A. Lucia, *Chemical Engineering Research and Design* **2014**, *92*, 2977–2991.

[155] M. R. Walsh, G. T. Beckham, C. A. Koh, E. D. Sloan, D. T. Wu, A. K. Sum, *The Journal of Physical Chemistry C* **2011**, *115*, 21241–21248.

[156] O. Konrad, T. Lankau, *The Journal of Physical Chemistry B* **2005**, *109*, 23596–23604.

[157] H Docherty, A Galindo, C Vega, E Sanz, *The Journal of chemical physics* **2006**, *125*, 074510.